

**TECHNOLÓGIAALAPÚ TESZTELÉS AZ OKTATÁSBAN:  
A PROBLÉMAMEGOLDÓ KÉPESSÉG FEJLŐDÉSÉNEK  
ÉRTÉKELÉSE**

Molnár Gyöngyvér  
Akadémiai Doktori Értekezés

2016

# TARTALOMJEGYZÉK

Tartalomjegyzék.....	1
Bevezetés.....	4
1. Az információs-kommunikációs technológiák hatása a tanulásra és oktatásra.....	7
1.1. A változtatás szükségessége az oktatásban .....	9
1.2. Az IKT oktatási integrációjának útjai.....	10
2. A papíralapú tesztek a számítógépes adaptív tesztesig: a pedagógiai mérés-értékelés technikájának fejlődési tendenciái.....	14
2.1. A kötött formátumú papíralapú tesztek és alkalmazási lehetőségeik .....	15
2.2. A kötetlen formátum és a valószínűségi tesztelmélet lehetőségei.....	17
2.2.1. Horgonyzási technikák.....	20
2.3. A valószínűségi tesztelmélet és a magyar alkalmazások.....	23
2.4. A technológiaalapú tesztelés különböző szintjei .....	25
2.4.1. A technológia szerepe a mérés-értékelés hatékonyságának növelésében .....	27
2.4.2. Az alkalmazott technológia típusának és módjának függvényében a technológiaalapú tesztelés típusai .....	28
2.4.3. Adaptív tesztelés .....	30
2.5. A technológiaalapú tesztelésre való áttérés előnyei, hátrányai és kihívásai.....	35
2.5.1. A tesztelés gazdaságossága, költségei.....	36
2.5.2. A közvetítés és adatáramlás gyorsasága, biztonsága .....	40
2.5.3. Azonnali, objektív, standardizált visszacsatolás – előnyök és kihívások .....	41
2.5.4. A tesztek jóságmutatóinak javulása .....	41
2.5.5. Innovatív feladatszerkesztési lehetőségek.....	42
2.5.6. Kontextuális adatok rögzítésének lehetősége.....	43
2.5.7. Adaptív tesztelés lehetősége.....	44
2.5.8. A tanulói motiváció változása .....	44
2.5.9. Bővül a tesztelésbe bevonhatók köre .....	45
2.6. Technológiaalapú mérés-értékelés kialakulása, fejlődése, nemzetközi és hazai tendenciái.....	46
2.6.1. Technológiaalapú mérések a nemzetközi összehasonlító felmérésekben .....	52
2.6.2. A számítógép-alapú tesztelés hazai fejlődése és tendenciái.....	56
2.6.3. A technológiaalapú mérés-értékeléssel kapcsolatos, a disszertációban felhasznált kutatásaink módszertanának áttekintése .....	58
3. A technológiaalapú tesztelés alkalmazhatósága kiskorúak körében .....	61

3.1. Az egér- és billentyűzethasználati képességek online mérése kisiskolás diákok körében .....	62
3.1.1. Célok .....	62
3.1.2. Módszerek .....	63
3.1.3. Eredmények.....	66
3.1.4. Az eredmények értékelése, további kutatási feladatok .....	73
3.2. Az iskolakezdés kulcsfontosságú előkészítéseinek online mérése kisiskolás korban .....	74
3.2.1. Célok, kutatási kérdések.....	75
3.2.2. Módszerek (minta, mérőeszköz, eljárások).....	75
3.2.3. A DIFER tesztek online változatával végzett mérések eredményei .....	82
3.2.4. A papíralapú, kisiskolás diákok részére kidolgozott induktív gondolkodás teszt online változatával végzett mérések eredményei .....	88
3.2.5. A technológiaalapú tesztelés alkalmazhatósága az iskolakezdés kulcsfontosságú előkészítéseinek mérésére: összegzés, az elemzések korlátai, továbblépési lehetőségek .....	90
4. A számítógép-alapú tesztelés alkalmazása 9-19 éves korban: a problémamegoldó képesség mérési lehetőségeinek változása .....	93
4.1. A problémamegoldó képesség mérési lehetőségei .....	93
4.1.1. A problémamegoldó képesség mérésének elméleti háttere és mérési tendenciái .....	94
4.1.2. A területspecifikus problémamegoldó képesség mérése papíralapú tesztekkel: a hazai kutatások szintetizálása.....	101
4.2. A problémamegoldó képesség számítógép-alapú mérése: célok és módszerek.....	115
4.2.1. Célok .....	116
4.2.2. Módszerek .....	118
4.3. A dinamikus problémamegoldó képességet mérő harmadik generációs tesztek működése: reliabilitás, faktorstruktúra, mérési invariancia, a problémák skálázása.....	125
4.3.1. A dinamikus problémamegoldó képesség tesztek jóságmutatói .....	125
4.3.2. A problémamegoldó képesség teszt hipotetizált kétdimenziós modelljének empirikus verifikációja .....	126
4.3.3. A problémamegoldó képesség teszt mérési invarianciája.....	128
4.3.4. A problémák skálázása: a nehézségi index változása az alkalmazott szcenárió függvényében.....	130
5. A dinamikus problémamegoldó képesség fejlődése és a fejlődést befolyásoló tényezők.....	138
5.1. A dinamikus problémamegoldó képesség fejlődése.....	138

5.2. A problémamegoldó képességszint kapcsolata az intelligencia és az induktív gondolkodás fejlettségi szintjével, valamint az Országos kompetenciamérés eredményeivel, az iskolai sikerességgel és demográfiai háttérváltozókkal.....	144
5.2.1. A problémamegoldó képességszint kapcsolata az intelligencia és az induktív gondolkodás fejlettségi szintjével.....	144
5.2.2. A problémamegoldó képességszint kapcsolata az Országos kompetenciamérés eredményeivel, az iskolai sikerességgel és demográfiai háttérváltozókkal.....	150
5.3. A gondolkodási képességek első generációs tesztekkel történő mérési eredményeinek előrejelző ereje a harmadik generációs teszttel mért problémamegoldó képességfejlettségi szintre .....	152
5.3.1. Az induktív gondolkodás, az intelligencia és a szülők iskolai végzettségének előrejelző ereje a diákok problémamegoldó képességszintjére .....	153
5.3.2. Az Országos kompetenciamérés, az iskolai sikeresség és a demográfiai háttérváltozók előrejelző ereje a diákok problémamegoldó képességszintjére .....	154
5.3.3. A diákok első évfolyamos DIFER, induktív gondolkodás, matematika és szövegértés teszteken nyújtott teljesítményének előrejelző hatása problémamegoldó képességük fejlettségi szintjére hetedik évfolyamon.....	155
5.3.4. A diákok nemének, hatodik és hetedik évfolyamos induktív gondolkodás, angol, matematika és szövegértés teszteken nyújtott teljesítményének előrejelző hatása problémamegoldó képességük fejlettségi szintjére hetedik évfolyamon.....	157
5.4. A dinamikus problémamegoldó környezetben alkalmazott felfedező stratégiák hatékonysága és azok változása: logfájl elemzések .....	160
5.4.1. A MicroDYN megközelítés teljes, alap és minimális stratégia modellje .....	161
5.4.2. A logfájlelemzések céljai .....	165
5.4.3. Módszerek .....	165
5.4.4. Eredmények.....	166
6. Összefoglalás és jövőbeli feladatok.....	187
7. Irodalomjegyzék .....	192
Köszönet.....	221
Rövidítések jegyzéke.....	222
Táblázatok jegyzéke.....	224
Ábrák jegyzéke.....	226
A disszertációban közvetlenül felhasznált saját közlemények jegyzéke .....	232
8. Mellékletek .....	234
8.1. Melléklet: Géphasználat teszt – 1. évfolyam, részlet (2014) .....	236
8.2. Melléklet: Dinamikus problémamegoldó képesség tesztek linkelése .....	247
8.3. Melléklet: Dinamikus problémamegoldó képesség teszt – részlet.....	248

## BEVEZETÉS

Az elmúlt néhány évtized legjelentősebb oktatási vonatkozású fejlesztései az oktatási rendszerek különböző szintjein megvalósuló visszacsatoló mechanizmusok kiépítésére irányultak, ezért óriási fejlődésen ment át a pedagógiai értékelés elmélete és gyakorlata. A világ vezető tesztközpontjai és szakemberei közreműködésével létrejöttek és rendszeressé váltak a nagy nemzetközi felmérések, amelyek hatására jelentős fejlődés történt az adatfelvételi technikák és technológiák, valamint az adatelemzési módszerek tekintetében. Fokozatosan kiépültek a nemzetközi trendeket és a nemzeti sajátosságokat is figyelembe vevő, nemzeti oktatási vonatkozású mérés-értékelési rendszerek, többek között a magyar Országos kompetenciamérés.

A gyors fejlődés következtében az ezredforduló idején leginkább elfogadott és elterjedt papíralapú tesztekre alapuló mérések egyre több korlátba ütköztek, sőt a papíralapú tesztekre alapozott fejlesztés lehetőségei mára teljesen kimerültek. A továbblépéshez, a 21. században jelentkező új mérés-értékelési igények kielégítéséhez alapvető, minőségi változtatásra volt szükség (Scheuermann és Pereira, 2008). A változtatás irányát jelentős mértékben meghatározta a technológia, a számítógépek fejlődése és oktatásban való terjedése, mindennapossá válása. A vonatkozó nemzetközi szinten jelentős kutatás-fejlesztési projektek (pl.: Assessment and Teaching of 21st Century Skills – ATC21S, Class of 2020 Action Plan; Griffin, McGaw és Care, 2012; SETDA, 2008) is, bár különböző oldalról közelítették meg a problémát, mégis arra a közös eredményre jutottak, hogy a továbblépés iránya kizárólagosan a számítógép-alapú tesztelésben rejlik (Scheuermann és Björnsson, 2009; Molnár, 2010a; Csapó, Ainley, Bennett, Latour és Law, 2012; Pearson, 2012). A számítógép-alapú értékelés ma már mind a hagyományos szemtől szemben történő, mind a papíralapú tesztelésnél hatékonyabb mérések megvalósítását teszi lehetővé, ezért belátható időn belül nagy valószínűség szerint megvalósul minden fontosabb mérés és vizsga technológiai alapokra helyezése. E tendencia mára már érzékelhetően megjelenik a nemzetközi szummatív, valamint a nemzeti segítő-fejlesztő, diagnosztikus mérési rendszerekben, fokozatosan bevezetve és tesztelve a számítógép-alapú adatfelvétel megvalósíthatóságát, előnyeit, lehetőségeit és a vizsgált konstruktumok esetleges változását.

Magyarország egyrészt részt vesz a fontosabb nemzetközi mérésekben, másrészt egy világszínvonalú nemzeti értékelési rendszert működtet. Mindezzel párhuzamosan elindult egy online segítő, fejlesztő, diagnosztikus mérési rendszer, az eDia kiépítése. A három értékelési rendszer bár teljes spektrumában (1-12. évfolyamig) átfogja a hazai közoktatást, ugyanakkor más-más céllal valósul meg. A nemzetközi értékelési rendszerek az országok oktatási rendszerét, a nemzeti szummatív értékelési rendszerek az iskolákat, az iskolák, pedagógusok hozzáadott értékét, az intézmény elszámoltathatóságát számszerűsítik a döntéshozók számára. A diagnosztikus értékelési rendszerek a diákok fejlődésének nyomon követését, az esetleges lemaradások időbeli jelzését valósítják meg.

Összességében a pszichológia kognitív forradalma, a 21. század gyorsan változó, állandó tanulás igénylő tudástársadalmában új tudáskonceptiók (pl. OECD, 2013c, 2013h, 2013i) kialakulása és a mérés-értékelés fejlődése – beleértve a hagyományos eszközökkel is mért területek életszerűbb mérési és fejlesztési lehetőségét, új konstruktumok mérhetőségét, motiváltabb, személyi képességszintjéhez illeszkedő tesztek kiközvetíthetőségét – lehetővé tette

annak meghatározását, hogy a fejlődés egyes szakaszaiban mit és hogyan érdemes mérni, valamint fejleszteni. Rávilágított és számszerűsítette az oktatás problémáit, valamint követhetővé tette az egyes oktatási beavatkozások hatását.

A hagyományos tesztelési technikákról a technológiaalapú tesztelésre történő átállás azonban nemcsak lehetőségeket, de számos kérdést, problémát és kihívást is állít a kutatók elé. A disszertáció középpontjában a számítógép-alapú tesztelésre való átállás szükségességének, előnyeinek és hátrányainak, valamint lehetőségeinek áttekintése áll, miközben a 21. században kulcsfontosságúnak tartott problémamegoldó képesség fejlettségi szintjét mérő kutatások szintetizálása segítségével konkrét példán keresztül ismertetem az adatfelvétel és adatelemzés fejlődési tendenciáit és a számítógép-alapú tesztelés tág életkori intervallumban történő megvalósíthatóságát.

A disszertáció öt fejezetből áll. Az első fejezetben áttekintem az információs és kommunikációs technológiák tanulásra és tanításra gyakorolt hatását, kitérek a változtatás oktatási szükségességére, valamint felvázolom az információs és kommunikációs technológiák (IKT) oktatási integrációjának útjait. Az egyes megközelítések értékelése során szem előtt tartom, hogy az eszközök, a módszertani repertoár és a mérés-értékelési kultúra változása csak akkor járul hozzá hatékonyan iskolai oktatásunk fejlesztéséhez, ha az oktatás valós problémáinak megoldásában alkalmazzuk azokat (*Csapó, Molnár és R. Tóth, 2009*).

A disszertáció második fejezete a pedagógiai mérés-értékelés technikáinak fejlődési tendenciáit ismerteti. Részletesen kitér a kötött és kötetlen formátumú tesztelésben lévő lehetőségek felvázolására, beleértve a valószínűségi tesztelmélet adta eszközrendszer innovativitását, ami a hagyományos értékelési technikákhoz képest jelentős mértékű fejlődést indukál a mérés pontossága, a kinyer információ mennyisége és a diák képességszintjéhez illeszkedő tesztelés megvalósíthatósága tekintetében. Hatékony alkalmazásának szükséges feltétele a technológia adatfelvételbe történő bevonása. A fejezetben részletesen ismertetem a technológiaalapú tesztelésre való átállás előnyeit, hátrányait és kihívásait, előbbi a mérés-értékelés minden egyes szintjén kimutatható (gazdaságosság, adatáramlás gyorsasága és biztonsága, visszacsatolás, tesztek jószágmutatóinak változása, feladatszerkesztési lehetősége, kontextuális adatok rögzítése, tanulói motiváció változása, a tesztelésbe bevonható diákok körének bővülése, személyre szabott tesztelés megvalósíthatósága). A fejezet utolsó egysége a technológiaalapú mérés-értékelés kialakulását, fejlődését, nemzetközi és hazai tendenciáit, illetve elterjesztésének feltételrendszerét tekinti át.

A harmadik fejezet fő kutatási kérdése, hogy a számítógép-alapú tesztelés alkalmazható-e kisiskolás korban. E célból áttekintem az egér- és billentyűzethasználati képességek, valamint az iskolakezdés kulcsfontosságú előkészítési feladatainak számítógép-alapú mérési lehetőségeit, miközben felvázolom e területek mérési tendenciáit a vizsgált konstruktum változását és ahol értelmezhető, a mérési invariancia meglétét vagy hiányát. A fejezetben bemutatott kutatások közös célja annak feltérképezése, hogy (1) kidolgozhatóak-e kisiskolás diákok körében alkalmazható, jó méréselméleti mutatókkal rendelkező számítógép-alapú tesztek, (2) hogyan viszonyulnak a számítógép-alapú teszteredmények korábbi hagyományos (szemtől szembeni vagy papíralapú) adatfelvételi technikákkal történő kutatások eredményeihez. A mintában lévő diákok fiatal életkora miatt nemzetközi szinten is hiánypótlóak a géphasználattal és az iskolakészültség vizsgálatával kapcsolatos kutatások.

Az egymásra épülő negyedik és ötödik fejezet a problémamegoldó képesség mérési lehetőségeinek változásán keresztül veszi górcső alá a számítógép-alapú tesztelésben rejlő mai

lehetőségeket, miközben tág életkori intervallumban (9-19 éves korban) ismerteti a diákok dinamikus problémamegoldó képességének fejlődését, valamint áttekinti a változás fő tendenciáit és lehetséges okait. A német és luxemburgi kollégákkal közösen fejlesztett, a számítógép-alapú tesztelés legújabb lehetőségeit is kihasználó, úgynevezett harmadik generációs tesztekkel kivitelezett kutatássorozat eredményei több szempont szerint is hiánypótlóak: tág életkori intervallumban vizsgálják a diákok problémamegoldó képességeinek fejlettségi szintjét és annak változását; ismertetnek egy, a diákok problémamegoldó stratégiáinak feltérképezését szolgáló logfájelemzésekhez kidolgozott modellt; a hét évet átfogó longitudinális adatok elemzésével azonosítanak több, a problémamegoldó képesség fejlettségi szintjét előrejelző változót; az adatfelvételek során mentett logfájlokon végzett strukturális modellelemzésekhez sorolható látens profil elemzések segítségével azonosítják a különböző típusú stratégiákat alkalmazó problémamegoldókat. Ezen elemzések segítségével olyan kutatási kérdésekre válaszolhattunk, amelyekre hagyományos technikákkal történő adatfelvételek és klasszikus tesztelméleti elemzések segítségével nem kerülhetett volna sor.

Szeged, 2016. március 8.

# 1. AZ INFORMÁCIÓS-KOMMUNIKÁCIÓS TECHNOLÓGIÁK HATÁSA A TANULÁSRA ÉS OKTATÁSRA

A 20. század ipari és 21. század tudás társadalma, gazdasága, kommunikációs szokásai, munkaformái, az értékesnek, versenyképesnek számító tudás jelentős mértékben eltér egymástól (*Trilling és Fadel, 2009*). A különbség egyik okát a technológia rohamos fejlődésében kereshetjük, ami néhány éven belül megváltoztatta az emberek szokásait, szórakozását, kapcsolattartási és vásárlási módjait, azt, ahogy dolgozunk, kommunikálunk, ügyeket intézünk, problémákat oldunk meg, és ismereteket szerzünk. Az interneten létrejöttek az érdeklődés alapján szerveződő fórumok, és rohamosan megnőtt a közösségi portálok tagjainak száma (*Csapó, 2008*). Az internet kiépülése és tartalommal való feltöltése alapvetően megváltoztatta a tudáshoz való viszonyunkat, a tudáshoz való hozzáférés lehetőségeit, költségeit (*OECD, 2010a, 2012*), sőt fokozatosan háttérbe szorította a többi információforrás szerepét.

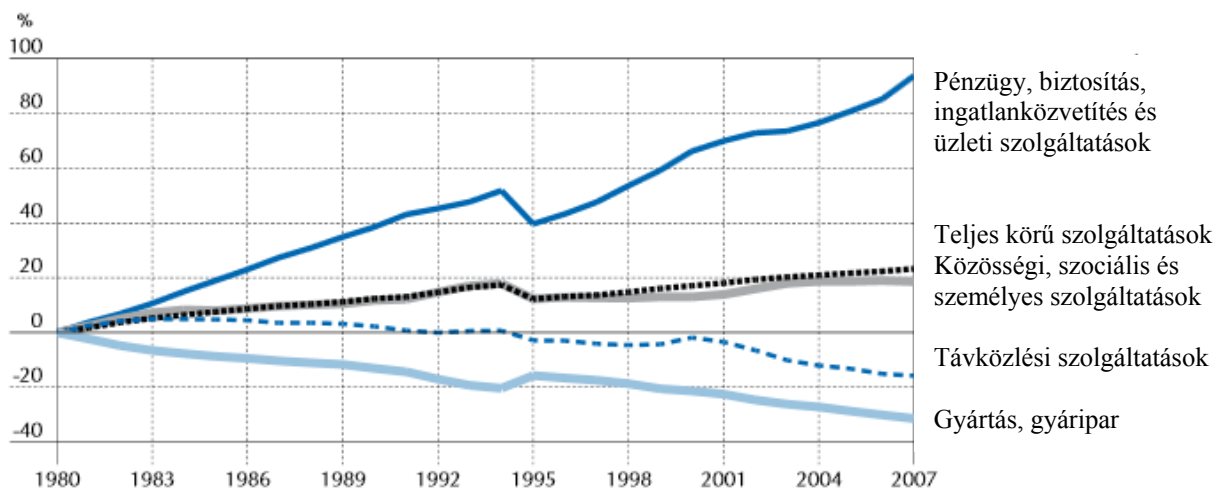
Az OECD (*Organisation for Economic Co-operation and Development*) 2013-as elemzése szerint az elmúlt négy évtizedben alapvetően megváltozott a munkahelyek jellege. Túlsúlyba kerültek a befektetési-, ingatlan-, biztosítási-, üzleti jellegű szolgáltatást biztosító munkahelyek száma, amelyek kivétel nélkül magasabb képességszintű emberek foglalkoztatására alapoznak (*OECD, 2013f*). A hatékony szolgáltatást nyújtó munkavégzés alapeleme az információhoz való gyors hozzáférés, azok hatékony elemzése és továbbítása, amely tevékenységek a 21. században elképzelhetetlenek technológiai eszközök és vonatkozó képességek (pl.: megfelelő szintű információs és kommunikációs technológiai képesség, kritikai gondolkodás, problémamegoldó képesség, kommunikációs képesség) birtoklása nélkül (1.1. ábra). Mindezzel párhuzamosan háttérbe szorultak az alacsony képzettséggel is ellátható, mechanikus munkafolyamatok, azok szerepét átvették a gépek, a technológia (*Trilling és Fadel, 2009*).

A technológia jelentős mértékben átalakította a társadalmi tőkét is (*Trilling és Fadel, 2009*). Jelenlétét társadalmi szinten befogadóvá teszi az, hogy közösen szerkesztünk, készítünk, és egymás között megosztunk médiatartalmakat (például: bloggolás, podcasting, Wikipedia, Flickr, YouTube), ismerősökkel és barátokkal erre alkalmas oldalakon vesszük fel, és tartjuk a kapcsolatot (Facebook, LinkedIn, Skype), továbbá használata nélkülözhetetlen a munkaerőpiacon. Mindezek következtében átalakult az értékesnek számító tudás, illetve képességek, kompetenciák köre, ami új kihívásokat állít a formális oktatás elé (*Goldman, Lawless, Pellegrino, Manning, Braasch és Gomez, 2012; Európa tanács, 2004; Law, Lee és Yuen, 2009*).

Az alkalmazható tudás, az új ismeretek létrehozását lehetővé tevő tudás, valamint az információs és kommunikációs technológiák (IKT) az ezredforduló meghatározó kifejezésévé váltak. Ennek hatására az oktatás, az oktatási rendszerek minősége, az oktatás során alkalmazott eszközök, taneszközök, módszerek ismét a figyelem központjába kerültek. Számos korábban értékesnek tartott ismeret, készség és képesség helyét új készség- és képességterületek vették át (*Mayrath, Clarke-Midura és Robinson, 2012a*). A 21. század gazdaságában, társadalmában, munkaerőpiacán a tények memorizálásának, egyszerű eljárások implementálásának már kevesebb szerep jut – ezek feladatát átveszik a különböző technológiai eszközök –, a hangsúly a flexibilitáson, a jó komplex problémamegoldó képességen, a hatékony kommunikációs



képességen és információkezelésen, a csoportmunkára való alkalmasságon, a kreatív és produktív technológiahasználaton, illetve az ezekkel összefüggő új tudás előállításának képességén van (Cisco, Intel és Microsoft, 2009; Quellmalz, Timms, Buckley, Davenport, Loveland és Silberglitt, 2012; Binkley, Erstad, Herman, Raizen, Ripley, Miller-Ricci és Rumble, 2012). Mindezen, a 21. században kulcsfontosságúnak számító készségeket és képességeket azonban ritkán tanítják a ma iskolájában (Graesser, 2012), holott nehéz elképzelni a jelen és jövő tanulási környezetét IKT eszközök és a vonatkozó készségek és képességek hatékony használata nélkül (RTTT, 2009; Obama, 2016).



1.1. ábra

Az ipari szektorban történő foglalkoztatás jellegének változása 1980 és 2007 között  
(forrás: OECD, 2013f. 48. o.)

Ezt támasztják alá a nemzetközi dokumentumokban olvasható oktatási vonatkozású célok is, melyekben a kezdeményezések ellenére jelentős mértékű változás nem történt az ezredforduló óta. Már a 2000-ben megfogalmazott *Lisszaboni célok* között is szerepelt, hogy Európa oktatását a tudásalapú társadalom elvárásainak fényében át kell alakítani úgy, hogy

- az emberi tőke fejlesztésével, a munkaerő produktivitásának fokozásával hozzájáruljon, segítse, támogassa a gazdasági növekedést;
- az IKT-s eszközök rutinszerű, hatékony használata segítse a tudásmegosztást, megkönnyítse az állami szolgáltatások ügyintézésének menetét (például: adóbevallás, Ügyfélkapu használata), a vásárlást, banki ügyintézés stb., illetve javítson a szociális kohézió;
- számítógép-alapú formatív és diagnosztikus mérés-értékelés bevezetésével hozzájáruljon az oktatás elszámoltathatóságának fejlesztéséhez, szummatív értékelés esetén pedig biztosítsa a hatékony fejlesztés alapját is jelentő gyors visszacsatolást (Kozma, 2008).

Az EU 2010-ig elérendő IKT-t érintő oktatáspolitikai ajánlásai között a következő főbb kritériumok fogalmazódtak meg:

- az IKT-eszközök jelenjenek meg az oktatás teljes területén, a képzés ne korlátozódjék felhasználói ismeretek átadására;
- az informatikai eszközök használata készségszinten épüljön be a tanítás-tanulás folyamatába;

- a tananyag tartalma és hozzáférhetősége legyen rugalmasabb, nyitott tanulási környezetben történjék az oktatás;
- az iskola alakítsa ki és fejlessze az élethosszig tartó tanuláshoz szükséges alapkészségeket, s készítsen fel az új tanulási formák és eszközök használatára.

Az előzőekben felsorolt *Lisszaboni célok* megvalósításának időpontja 2010 volt, de a nem teljesítés (l. pl.: *OECD*, 2013g) következtében közel azonos oktatási vonatkozású célok fogalmazódtak meg a Lisszaboni célok II (Európa 2020; *World Economic Forum*, 2010) dokumentumokban is. A cél továbbra is a magas színvonalú formális oktatás és tréning megvalósítása maradt, ami biztosítja, hogy a tagországok képesek legyenek alkalmazkodni a gyorsan változó gazdasági környezethez, ami segíti a munkaerő-piaci viszonyok javulását (*World Economic Forum*, 2010).

### 1.1. A változtatás szükségessége az oktatásban

Néhány évtizede beszélhetünk internetről, 1977 a személyi számítógép megjelenése, 1989 a www debütálása – új elvárások az IKT és az oktatás kapcsolatát tekintve –, 1998 a nyílt forráskódú programok megjelenése, 2001 a dotcomösszeomlás, majd 2005 a web 2.0, az írható web, ami alapjaiban változtatta meg az internetezési és ezzel párhuzamosan a tanulási szokásokat (*White*, 2008). Mindennek következtében az IKT nem rendelkezik több évszázadra visszatekintő oktatási tradícióval, mégis jelentős, minőségi változtatást generáló erővel bír.

A 21. század és a 20. század diákjai között olyan mértékű különbség van, ami nemcsak, hogy új igényeket támaszt az oktatás felé, de ezek az új igények már nem kivitelezhetők a hagyományos módszerekkel. Ennek hatására kialakult az igény egy új típusú iskola iránt. Az ezredforduló óta az oktatási rendszerek egyre növekvő nyomás alá kerültek és kerülnek, hogy integrálják a különféle technológiai eszközöket, továbbá azok segítségével lehetőséget biztosítsanak a diákoknak a 21. században kulcsfontosságúnak tartott képességek elsajátítására (*UNESCO*, 2002), és úgy alakítsák át mérés-értékelési rendszerüket, hogy az alkalmas legyen ezen új tudás vizsgálatára is. Nem kétséges, hogy a technológia adta új lehetőségeknek megvan az a képessége, hogy alapjaiban megváltoztassa, transzformálja az oktatást. Ennek támogatására az utóbbi néhány évben számos nagy nemzetközi szervezet (*OECD*, *UNESCO*) kutatásokat indított.

Annak ellenére, hogy az ezredforduló óta számos kutatási jelentésben, oktatáspolitikai dokumentumban szerepeltek az előzőekben felsorolt célok, az országok jelentős részében, így hazánkban sem történt jelentős mértékű minőségi változás. Az IKT oktatási megjelenését mérő, értékelő, jellemző kutatások fókusza ezekben a térségekben még az ezredforduló után tíz évvel is az alkalmazott eszközök mennyiségére, az alkalmazás gyakoriságára, típusára és az eszközhasználati szokásokra kérdez rá (*Molnár és Kárpáti*, 2012; *Molnár és Pásztor-Kovács*, 2015; *Hunya, Dancsó és Tartsayné*, 2006; *Hunya*, 2011, 2013a, 2013 b, 2015; *Law, Pelgrum és Plomp*, 2008; *European Commission*, 2013; *Wastiau, Blamire, Kearney, Quittre, Van der Gaer és Monseur*, 2013). Ezek az Európai Unió dokumentumaiban is általánosan használt indikátorok azonban kevésbé alkalmasak a technológia oktatási integrációja szintjének átfogó és alapos leírására, jellemzésére. A kutatási eredmények hiánya ellenére azonban általánosan elfogadott, hogy az IKT fontos szerepet játszik az oktatás és tanulás átalakításában, átalakulásában, modernizálásában, a 21. században már senki sem kételkedik az IKT oktatásban betöltött szerepének fontosságában (*Molnár és Magyar*, 2015).

Vannak országok, ahol a társadalom IDI indexe (*ICT Development Index* – IKT fejlődési index) jelentős (Korea, Dánia, Izland, Anglia, Svédország, Finnország, Norvégia, Hong Kong; l. *ITU*, 2015), vagy jelentős IDI indexnövekedésen esett át (pl.: Thaiföld, Kazahsztán; l. *ITU*, 2015), ahol az IKT az oktatás-tanulás szerves részévé is vált (l. pl.: *KCC*, 2014). Ezen országok, nemzetközi oktatási eredményeik (*OECD*, 2013a, 2013b) alapján, mind az OECD PISA mérések élvonalában helyezkednek el (pl.: Finnország, Korea), vagy a jelentős fejlődést elért országok között vannak (l. Malajzia, Kazahsztán). A PISA méréseken még Finnországot is jelentős mértékben megelőző Dél-Koreában (bármely területet tekintve) ma már kizárólagosan számítógépen tanulnak a diákok, a papíralapú tankönyvek megszűntek, megvalósítva az IKT teljes mértékű oktatási integrációját.

## 1.2. Az IKT oktatási integrációjának útjai

Az IKT eszközök iskolai elterjesztésének módja eltérő irányú folyamatokat indíthat el. Egyrészt a technológia iskolai integrációja számos új lehetőséget kínál, másrészt társadalmi csoportok és egyes személyek között akár tovább növelheti a fennálló eltéréseket. Míg egy jól átgondolt fejlesztés csökkentheti az iskolák között lévő különbségeket, addig, ha ugyanez a folyamat spontán zajlik, a jobb érdekérvényesítő iskolák még nagyobb előnyre tesznek szert. Ezzel növelve az iskolák közötti különbség nagyságát és ezáltal tovább szélesítve a tanárok, diákok, iskolák között fennálló „digitális szakadékot”. A technológia használata ugyanis több lehetőséget kínál, mint bármely korábbi taneszköz. Nemcsak a többcsatornás ismeretközlést és ezáltal a tudás új reprezentációs formáinak kialakítását teszi lehetővé, hanem motiváló erővel is bír a diákok irányában. A megfelelő szintű IKT használat számos egyéb tantárgyi kompetencia katalizátora is. Empirikusan bizonyított (*Pelgrum*, 2004), hogy akik rendszeresen használják a számítógépet, átlagosan magasabb az olvasási képességszintjük, sőt, motiváltabbak a természettudományok elsajátítására is. Azok a diákok, akik rendszeres digitális szöveg olvasók, azaz gyakran olvasnak e-maileket, chattelnek, online híreket olvasnak, online szótárakat, enciklopédiákat használnak, részt vesznek online fórumbeszélgetéseken, információt keresnek Interneten, általában véve magasabb képességszintű olvasók, mint akik keveset olvasnak online (*OECD*, 2010a, 2010b) – azonos mennyiségű papíralapú olvasást feltételezve.

A technológia és információs tudástársadalom adta lehetőségek kihasználásához több feltétel együttes teljesülése is szükséges, azok egymagukban nem oldják meg a problémákat. A technológia tanításba, tanulásba történő integrálását több oldalról támogathatjuk mindenképp szem előtt tartva, hogy ne a technológia határozza meg a változtatások irányát, az a változtatások katalizátora legyen.

A technológia oktatási integrációjára való törekvés három fő iránnyal (*Selwyn*, 2011), céllal jellemezhető: 1) virtuális iskolák alakítása, 2) digitálisan vezényelt újraiskolázás (minőségileg más oktatás, értékelés), 3) a társadalom iskoláztatása (otthonról tanulás, a mai értelemben vett formális tanulás megszüntetése). Egy nemzet kötelező iskoláztatása véleményem szerint sem az első, sem a harmadik úton nem valósítható meg. A második út kapcsán is számos megközelítéssel találkozhattunk, melyek egyértelműen az előző részben kifejtett digitális változtatásban látják és látták a megoldást.

A továbbiakban felsorolt lehetőségek külön-külön a hatékony integráció szükséges, de nem elégséges feltételei:

- 1) mind a formális (iskolai), mind az informális (otthoni) tanulás során megfelelő felszerelés (hardver és szoftver) biztosítása – jelen pillanatban nagyon nagy különbségek vannak az egyes iskolák felszereltsége között (*Molnár és Pásztor-Kovács, 2015a*).
- 2) Nemzeti tantervben való megjelenés – az eszközök használatának készségei és képességei elsajátítását túllépve az egyszerű technikai készségeket lehetővé téve a virtuális világban való eligazodást, a szükséges információ gyors és hatékony megtalálását tanítani.
- 3) Tanárképzésbe történő integráció – nem elegendő az eszközök használatának megtanítása, az IKT-s képességek fejlesztése, a különböző webes alkalmazások megismerése, multimédiás anyagok magas színvonalú készítése –, holott utóbbi elkerülhetetlen feltétele az IKT-ban rejlő interaktív potenciál tanítási-tanulási folyamatban történő kiaknázásának – filozófia, nézet, tanításról kialakított kép, pedagógiai megközelítés, módszertani repertoár, mérés-értékelési módszertan váltása, tanár és diákszerep megváltoztatásáról van szó, ami lehetővé teszi, hogy a tanárok az IKT-t, mint pedagógiai eszközt és ne célt sajátítsák el. A technológia a diákok közötti kollaborációt, a közösségi szinten történő aktív kommunikációt és a hatékonyabb tudásépítést és tudásmegosztást támogassa (*Molnár, 2011a*).
- 4) A mérésekben történő megjelenés (pl.: Országos kompetenciamérés, érettségi) – a nemzetközi empirikus vizsgálatok fokozatosan átállnak a papíralapú tesztelésről a számítógép-alapú tesztelésre (2015-től a PISA tesztek is kizárólag számítógép-alapúak lesznek), ami gyorsabbá, pontosabbá és hatékonyabbá teszi a mérés-értékelés egész folyamatát, sőt olyan készségek mérésére ad lehetőséget, amire papíralapon nincs lehetőség (*Csapó, Molnár és R. Tóth, 2008; Molnár, 2010a*). Ezt a tendenciát, a papíralapú tesztelésről a számítógép-alapú tesztelésre való fokozatos átállást javallott hazánkban is követni és meg kell teremteni annak feltételeit.

Az ezredforduló környékén számos ország jelentős beruházást indított el a technológia oktatásba való integrálása, a technológia oktatásban betöltött szerepének növelése és javítása érdekében. Ennek hatására a PISA-adatok 2000 és 2003 között nemzetközi szinten exponenciális növekedést detektáltak az IKT oktatásban történő jelenlétét illetően (*OECD, 2006*). Három év alatt jelentős mértékben javult az egy diákra jutó számítógép és internetes kapcsolat aránya, azonban a befektetések főképp hardveres fejlesztések voltak. Ez a jelenség hazánkban is tapasztalható volt, 2000 és 2004 között mind a magánéletben, mind a munkahelyen elvárassá vált a számítógép és internethasználat (*ITTK, 2007*).

A formális tanulást támogató eszközbeszerzések tekintetében Magyarországon nemzetközi szinten is korainak volt nevezhető a Sulinet program, ami 1997-ben indult azzal a céllal, hogy internetre csatlakoztassa és számítógépekkel szerelje fel az iskolákat, majd a géppark fenntarthatóságát, fejlesztését biztosítandóan 2005-ben a Közoktatási informatikai fejlesztési program. Nagy visszhangot keltett a digitális zsúrkocsi és bőrrönd, továbbá az iskolák interaktív táblához juttatását megcélzó program. Utóbbi célja volt, hogy 2010-re az osztályok felét hardveresen felszerelje (interaktív tábla, projektor és számítógép). Szintén 2010-ig kitűzött cél volt az egy diákra jutó számítógépek számának arányán tovább javítani (1:6 arány elérése), továbbá a tanárok felét saját notebookhoz juttatni. Mindezek ellenére 2006-ban a tantermek 19%-a volt csak számítógépekkel felszerelve (EU-átlag 68%), illetve a tanórák kb. 3%-ban kapott helyet a technológia, ott is a korábban alkalmazott oktatási módszer támogatására használták a számítógépet. Az Európai Bizottság 2000-res javaslata szerint az általános iskolában minden nyolc, a középiskolában minden hat diákra kellene, hogy egy számítógép

jusson. A hazai adatokat tekintve ez az arány 2009-ben (a számolásból kizárva az elavult és működésképtelen számítógépeket) 1:27-hez (Kárpáti és Horváth, 2009); 2011 elején hazai reprezentatív minta alapján az általános iskolák vonatkozásában 1:15-höz volt. Ha kizárjuk az elemzésből a hat év és annál idősebb, azaz elavult számítógépeket, 1:19-hez (Tóth, Molnár és Csapó, 2011), 2014-ben pedig általános iskolákra 1:9, középiskolák esetén pedig 1:6 volt ez az arány (Molnár és Pásztor-Kovács, 2015a). Az iskolák infrastrukturális helyzetét országos reprezentatív mintán feltérképező empirikus kutatás eredményei alapján megállapítható, hogy a 2010-re előírt indikátorszámokat 2014-re sikerült az általános iskolák esetében megközelítenünk, a középiskolák kapcsán elérnünk (l. Molnár és Pásztor-Kovács, 2015a).

A nemzetközi viszonylatban relatív alacsony technológiahasználat egyik oka a tanárok technológiához való hozzáállása volt. 2006-ban még a tanárok 96%-a gondolta úgy, hogy az informatika oktatása külön tantárgy keretein belül történjék (EU-átlag 54%), és csak 36% mondta azt, hogy más tantárgy keretein belül is tanítani kellene az informatikai műveltséget (EU-átlag 76%). Az általános iskolai tanárok 32%-a szerint felesleges ezen eszközök használata (EU-átlag 3%). Az ezredforduló után 10 évvel a TÁMOP- és TIOP-pályázatok nyújtottak lehetőséget a technológiaalapú iskolai fejlesztésekhez. A kormány az informális tanulás, azaz az otthoni géphasználat terjedését hivatott segíteni a 2006-ban zárult, sikerességét sokat vitatott Sulinet Express programmal.

A legtöbb esetben a felülről lefelé irányuló „top-down” koncepció volt megfigyelhető, azaz építsük ki a technológiát, szereljük fel az iskolákat, majd az integráció, az eszközök adaptációja jön magától „build it and they will come” (Scheuermann és Pedró, 2009). Feltételezték, hogy mind a tanárok, mind a diákok egy idő után elkezdik kreatívan használni a technológiát. Ebből adódóan a legtöbb országban gyakran felmerülő kérdés, hogy mikor fognak ezek a befektetések megtérülni, vajon a technológia jelenléte beteljesíti-e a hozzá fűzött elvárásokat?

Az elvárás, miszerint, ahogy bekerül a technológia a tanterembe, a tanárok elkezdik produktívan használni, ezáltal jelentős mértékben átalakul a tanítási és tanulási folyamat, nem volt teljesen realiztikus, amit a magyar tanárok IKT-hoz való viszonyával kapcsolatos számok is mutatnak. Ennek ellenére ez a lépés, a technikai feltételek megteremtése, a környezet biztosítása elkerülhetetlen volt. A hosszabb távú hatékony használat feltétele azonban az is, hogy az elavult eszközöket lecseréljék, a programokat frissítsék, ami ismét plusz költséget jelent a kormányzatok számára. Ennek biztosítása a kezdeti lelkesedés lankadása után már több ország esetében problémás.

Az integráció segítésének egy lényeges pontja, ha az IKT megjelenik a különböző nemzetek tanterveiben, amik meghatározzák, hogy a diákok hogyan, illetve mire használják a különböző technológiai eszközöket a formális oktatás keretein belül. Hazánkban külön informatika órán tanulják a diákok az eszközök használatát, ennek fényében a fent említett eszközpark is általában elzárt számítástechnika termekben található, és a kereszttantervi kompetenciák tanítását célzó IKT-integráció csak abban az esetben valósulhat meg, ha az adott óra a fent említett termekben zajlik. Központilag minden egyes tantárgy tekintetében a technológiaalapú tanítást támogatja és segíti a 2004-ben indult tananyag-adatbázis, a Sulinet Digitális Tudásbázis (SDT; <http://sdt.sulinet.hu>) program. Nemzetközi viszonylatban több országban nincs külön informatika óra, hanem a többi tantárgy keretein belül sajátítják el a diákok a hatékony IKT használatához szükséges ismereteket, képességeket, kompetenciákat (EETT, online).

Az eszközök használatának szükséges feltétele, hogy az megjelenjen a tanárképzésben, illetve a tanártovábbképzésben is. Ez nem kis kihívás elé állítja a tanárképzést, miután az IKT-s eszközök hatékony integrációjához nem elegendő az eszközök használatának megtanítása, az IKT-s képességek fejlesztése. Még az sem elegendő, ha a tanárok, illetve leendő tanárok megismerik a különböző webes alkalmazásokat, megtanulják, hogyan alkalmazhatók azok a mindennapi tanítás során, illetve gyakorlattá válnak multimédiás anyagok magas színvonalú készítésében. A folyamatot nehezíti a technológia rapid fejlődése, változása, aminek hatására nemcsak egy eszköz használati módját, hanem egy új típusú gondolkodásmódot, nyitottságot is el kellene sajátítani. A technológia gyors változása következtében valószínű, hogy amire az adott eszköz, szoftver élesben használatra kerül, már jelentős mértékben átalakul a fejlesztések következtében. Az előzőek megvalósításához azonban jelentős mértékű szerepváltásra van szükség mind a tanár, mind a diák oldaláról nézve.

Az IKT iskolai alkalmazása, oktatásba történő bevonása nemcsak az újabb eszközök tantermi megjelenésével és azzal párhuzamosan a tanárok IKT-s módszertani repertoárjának növelésével lehetséges, hanem az iskolai mérés-értékelés folyamatába történő integrálással is. Mindez csak akkor járul hozzá hatékonyan az oktatás fejlesztését, ha nem a megjelenő technikákhoz keressük a felhasználás lehetőségeit, azaz nem a technológia a cél, hanem az oktatásban jelentkező problémák hatékony megoldására alkalmazzuk azokat (*Csapó, Molnár és R. Tóth, 2009*).

## 2. A PAPÍRALAPÚ TESZTEKTŐL A SZÁMÍTÓGÉPES ADAPTÍV TESZTELÉSIG: A PEDAGÓGIAI MÉRÉS- ÉRTÉKELÉS TECHNIKÁJÁNAK FEJLŐDÉSI TENDENCIÁI

Az elmúlt másfél évtized egyik legdinamikusabban fejlődő területe a pedagógiai mérés-értékelés. Az ezredforduló óta mind hazai, mind nemzetközi szinten kiépültek és jelentős mértékű fejlesztésen, illetve fejlődésen mentek keresztül a rendszerszintű és intézményszintű értékelést megvalósító értékelési rendszerek (pl.: OECD PISA, IEA PIRLS, NAEP, Kompetenciavizsgálat; R. Tóth, Molnár, Latour és Csapó, 2011). Az ezredforduló idején leginkább elfogadott és elterjedt papíralapú tesztekre alapuló mérések a fejlesztések ellenére egyre több korlátba ütköztek, a papíralapú tesztekre alapozott fejlesztés lehetőségei mára teljesen kimerültek.

A továbbblépéshez, a 21. században jelentkező új mérés-értékelési igények kielégítéséhez alapvető, minőségi változtatásra van szükség (Scheuermann és Pereira, 2008). Ez a felismerés nemzetközi szinten jelentős kutatás-fejlesztési projekteket indukált (pl.: ATCS21, Class of 2020 Action Plan; Griffin, McGaw és Care, 2012; SETDA, 2008), melyek egyöntetűen a számítógép-alapú tesztelésre való áttérésben jelölték meg a továbbblépés irányát (Scheuermann és Björnsson, 2009; Molnár, 2010a; Csapó, Ainley, Bennett, Latour és Law, 2012; Pearson, 2012; Pearson Educational Measurement, 2013). E kutatási eredmények hatása, a számítógép-alapú tesztelésre való átállás javaslata mára már érzékelhetően megjelenik a nemzetközi mérési rendszerekben, fokozatosan bevezetve a számítógép-alapú tesztelést.

Az ismertebb OECD PISA felmérésekben például fokozatosan egyre több területen és egyre több ország részvételével került bevezetésre a számítógép-alapú tesztelés, először a papíralapú tesztekkel párhuzamosan, majd 2015-től kizárólag számítógép-alapú tesztekkel oldanak meg a diákok.

Ha biztosítani szeretnénk a továbbfejlődés lehetőségét, szükséges, hogy (1) a technológiaalapú mérés-értékelés fokozatos bevezetésével segítsük az iskolák mérési-értékelési kultúrájának továbbfejlődését, (2) megismertessük a pedagógusokkal a modern mérési eszközöket, (3) biztosítsuk az iskolák és iskolafenntartók számára azokat az adatokat és eljárásokat, amelyekkel intézményük/ intézményeik objektív értékelését el tudják végezni és azt megbízhatóan össze tudják hasonlítani az országos adatokkal. Az áttérés iskolai kontextusban – mint tapasztaljuk a PISA mérések ez irányú változtatásai kapcsán – azonban csak fokozatosan lehetséges, minden lépésben gondosan ellenőrizve, és kiszűrve a nemkívánatos mellékhatásokat (Csapó, Molnár és R. Tóth, 2008).

Az értekezés e fejezetében (1) a kötött formátumú, hagyományos, papíralapú tesztek főbb jellemzőinek fényében ismertetjük a kötetlen formátum és a valószínűségi tesztelmélet adta lehetőségeket, (2) áttekintjük a technológiaalapú tesztelés különböző szintjeit, előnyeit, hátrányait és kihívásait, valamint (3) a pedagógiai mérés-értékeléssel foglalkozó főbb (reprezentatív mintán alapuló) kutatások hazai megjelenésének és irányvonalainak fényében bemutatjuk a technológiaalapú mérés-értékelés hazai és nemzetközi tendenciáit.

## 2.1. A kötött formátumú papíralapú tesztek és alkalmazási lehetőségeik

A számítógépes tesztelés sajátosságainak ismertetése előtt összefoglaljunk a hagyományos, papíralapú tesztek jellemzőit, ugyanis ezekhez viszonyítva lehet megmutatni azokat az új lehetőségeket, amelyeket a számítógépes tesztelés kínál, és azokat a problémákat, kihívásokat, amelyeket az új mérési technikák felvetnek. Az úgynevezett hagyományos, közismert, papíralapú (*Paper-and-pencil* – PP) tesztek nagyon fontos szerepet játszottak és játszanak ma is a tanítási-tanulási folyamatok irányításában, az oktatás eredményességének felmérésében. Ezek a tesztek többnyire rögzített formátumúak (*Fixed Form* – FF), ami azt jelenti, hogy a tesztek feladatait mindig azonos formai elrendezésben kapják meg a tesztelt személyek. Szigorú értelemben csak így biztosítható a teszt objektivitása, azaz, hogy az mindig mindenkit egyformán mér. Az item pozíciós hatását vizsgáló kutatási eredmények (*Magyar és Molnár, 2015; Hahne, 2008*) szerint ugyanis a feladatok sorrendjének szerepe lehet, van a megoldás valószínűségében.

A PP FF tesztekben sokféle item (a legkisebb, önállóan értékelhető egység) fordulhat elő, változatos item-formátumokat használhatnak, ezek csoportosításának egyik dimenziója a zárt-nyitott kérdéstechnika. A zárt, vagy feleletválasztós kérdések esetében előre megadott válaszokból választva kell a tesztet megoldani. Az ilyen feladatokból álló teszteket gyakran nevezik objektív teszteknek, mivel azok értékelése nem igényel személyes emberi döntéseket. A leggyakrabban alkalmazott objektív item-formátumok a többszörös választás (*multiple-choice*) és a dichotóm választás (alternatív választás, tekinthető a többszörös választás speciális esetének), amelynek egyik formája az igaz-hamis döntés (*true-false*). Ugyancsak objektív item-formátum az illesztés (párosítás, *matching*), melynek során két halmaz elemei között kell megfeleltetést létrehozni.

A nyitott, vagy feleletalkotó (*Constructed Response*, CR) kérdések esetében a tesztelt személy maga alkotja meg a választ, aminek értékelése, a válasz helyességének megállapítása további, többnyire személyes kódolói döntést igényel. A CR itemek az objektivitás szempontjából egy szélesebb spektrumot alkotnak a rövid választól (egy kifejezés, egy szó vagy egy szám a válasz) az esszé jellegű kérdésekig. Attól függően, hogy mennyire sokféle lehet a válasz, az értékelő (kódoló) lehetőségei is bővülnek. Így már csak bizonyos közelítéssel biztosítható, hogy egymástól független értékelők ugyanolyan módon döntsének egy válasz helyességét illetően. A CR tesztek objektivitását az egyértelmű javítókulccsal, kódolási utasítással és az értékelők képzésével lehet javítani.

A zárt és a nyitott tesztfeladatok alkalmazása közötti választás során két ellentétes szempontot kell mérlegelni. Egyrészt az objektív itemek – mivel nem igényelnek további emberi értékelő beavatkozást – olcsóbbak, gyorsabban lehet az eredményekhez jutni. Megválaszolásuk a teszt megoldójától is kevesebb időt igényel, a kész válaszok közötti döntés gyorsabb lehet, mint a válasz önálló megalkotása. Éppen ebből következően másfajta gondolkodást igényel(het)nek, mint az önálló válaszadás, ezért csak a tudás bizonyos komponenseinek mérésére alkalmasak. A CR itemek – ha azok kódolása emberi munkával történik – kevésbé objektívek, feldolgozásuk drágább és lassúbb, viszont a tudás változatosabb formáinak felmérésére alkalmasak.

A PP FF tesztek készítésének és fejlesztésének alapjául hosszú időn keresztül a klasszikus tesztelmélet szolgált (bővebben l. pl. *Csapó, 2000*). Ez egy szigorú, axiomatikus matematikai elmélet, amelynek következtetései alkalmasak a tesztek minőségének jellemzésére. Az elmélet



alapvető feltevése szerint minden felmért személy rendelkezik a vizsgált tulajdonság egy  $V$  valódi értékével, és minden mérés szolgáltat róla egy  $M$  mért értéket. A két érték közötti különbség a hiba, korrelációjuk pedig a teszt megbízhatóságát, reliabilitását jellemző mutató. Mivel a  $V$  közvetlenül soha nem határozható meg, az említett korrelációt sem lehet közvetlenül kiszámítani. A klasszikus tesztelmélet tételeit felhasználva azonban bizonyos mérhető adatokból lehet arra becslést adni. Például a megismételt tesztelés adataiból, vagy a teszt belső konzisztenciájából (az itemek működésének összefüggéseiből). Az egyes itemek minőségét is a teszthez képest lehet megítélni: más itemekkel, főleg pedig a teszt-összpontszámmal való korreláció jól megmutatja, illik-e egy item a tesztbe, ugyanazt méri-e, mint a többi.

A tesztek elemzésének, a hibás, rosszul mérő itemek kiszűrésének, az itemek fejlesztésének a klasszikus tesztelméletre épülő kifinomult technikai alakultak ki, és az egymást követő kipróbálás és javítás után nagyon jó minőségű tesztek lehet készíteni. A fejlesztés eredményeként matematikailag akkor nő a reliabilitás, ha a teszt homogén, egymással magasan korreláló, és közepes nehézségű itemekből áll. Ez az oktatási alkalmazások szempontjából nem mindig előnyös, mert fontos mérendő tartalmak szorulhatnak így ki a tesztből. A közepes nehézség pedig azzal járhat, hogy az átlagostól felfele vagy lefele eltérő teljesítmények mérésére a teszt kevésbé alkalmas.

A PP tesztek felbontása, azaz, hogy egymáshoz mennyire közel álló teljesítményeket lehet velük megkülönböztetni, meglehetősen korlátozott. Ha például egy teszt 20 itemből áll és minden egyes item megoldásával 0 vagy 1 pontot lehet elérni, akkor az egymástól 5% távolságra levő teljesítményeket lehet csak az adott teszttel megkülönböztetni. A felbontást az itemek (elméleti vagy tapasztalati) súlyozásával lehet finomítani, azonban a kötött formátum mellett, ha mindenki ugyanazokat a feladatokat oldja meg, a felbontás javításának komoly korlátjai vannak.

A PP FF tesztekkel az említett korlátokból fakadóan csak egy viszonylag szűk képességtartományt lehet jól felmérni. Ha a teszt egy szélesebb képességtartományt fog át, akkor minden egyes felmért személynek csak a feladatok egy szűkebb sávja jelent valódi kihívást, amely a saját képességéhez közel álló feladatokat tartalmaz. A feladatok nagyobb része viszont vagy túlságosan könnyű, ezért unalmas, vagy túl nehéz, ezért frusztráló hatású lehet. Egy-egy alkalommal elvégzett tesztelésnél ezek a hatások nem túl jelentősek, ha azonban az oktatási folyamatba rendszeres tesztelés épül be, az említett negatívumok már komolyan veszélyeztetik az érdeklődést, a teszteléssel kapcsolatos attitűdöt és a feladatok megoldásához szükséges motivációt.

A tesztek alkalmazásának egy további jellemzője, hogy mekkora tétje van a teszteredménynek a felmért egyén számára. Ebből a szempontból megkülönböztethetjük az alacsony téttel (*low stakes*) és a magas téttel (*high stakes*) megoldott tesztek. Ez tehát nem magának a tesztnek, hanem a tesztelés kontextusának a jellemzője. Például az érettségi vizsgának kifejezetten magas a tétje, de a próba-érettséginek elhanyagolható. Természetesen az alacsony vagy magas tét csak a két végpont megnevezése, hiszen a tét nagyságát tekintve itt is egy folytonos változóról van szó. Mindez alapvetően befolyásolja a tesztmegoldók motivációját, érdekltségét, és késztetését a mérés céljaitól idegen módszerek és eszközök alkalmazására. Például a tesztmegoldások betanulása, illegális segédeszközök alkalmazása annál valószínűbb, minél nagyobb a tesztelés tétje. A teszt alkalmazóinak ezzel arányos erőfeszítéseket kell tenniük a tesztelés objektivitásának biztosítása, például a feladatok titokban tartása érdekében (Csapó, 2014).

Ez utóbbi szempontok úgy függenek össze a tesztek formátumával és minőségével, hogy a tesztek – az előbb említett reliabilitási problémák miatt is – többszörösen ki kell próbálni, a nem jól mérő itemeket szükség esetén korrigálni kell. Amíg azonban a formatív tesztek nyilvánosan lehet kezelni, folyamatosan lehet fejleszteni, és alkalmazni, a magas tétellel bíró kontextusban alkalmazott kötött formátumú tesztek titkosan kell kezelni, és többnyire csak egyszer lehet alkalmazni. Ebből következik az a paradoxon, hogy minél nagyobb egy kötött formátumú teszt tétje, annál nehezebb azt kipróbálni, fejleszteni, javítani. Ez azonban nem adhat felmentést arra, hogy tömegével alkalmazzanak fiatalok sorsát eldöntő, ugyanakkor megkérdőjelezhető minőségű tesztek. A kipróbálásnak ebben az esetben is meg lehet találni a módszereit, bár azok nyilvánvalóan költségesek.

## 2.2. A kötetlen formátum és a valószínűségi tesztelmélet lehetőségei

Az oktatási kontextusban alkalmazott mérések többnyire nem egyetlen kötött formátumú tesztet igényelnek, mert például olyan nagy tudásterületet vizsgálnak, vagy olyan széles képességfejlődési spektrumot kellene átfogniuk, amelyek technikai okokból sem férnek bele egyetlen tesztbe. A probléma megoldására számos technika született. Ezek közé tartozik a teljes lefedés elve, amikor egy nagyobb tudásterület teljes felméréséhez a lehetséges összes feladat elkészül. Ilyen megoldást dolgozott ki *Nagy József* az általa irányított program elméleti keretével, amikor a fontosabb iskolai tárgyak teljes tudásanyagát magában foglaló tesztek készültek (*Nagy, 1972*). Ilyen esetben az elkészült feladatokat ekvivalens tesztváltozatokba sorolják úgy, hogy minden egyes tesztváltozat kezelhető méretű legyen. Így, bár az országos reprezentatív felmérések során egy tanuló mindig csak az összes feladat egy részét oldotta meg, a felmérés egészéből az összes tudáselem elsajátításáról képet lehetett alkotni.

Egy másik megoldás a feladatbankok alkalmazása, amikor lényegében a teljes lefedés előzőekben bemutatott elveit alkalmazva, tesztváltozatokba sorolva kerül sor a feladatok bemérésére. Ezután az összes feladat egy feladatbankot alkot, amelyből a konkrét felmérések igényeinek megfelelően lehet kiválasztással vagy véletlen sorsolással a konkrét felmérések céljaira tesztek összeállítani. Erre a megoldásra is lehet egy korai példát bemutatni a magyarországi gyakorlatból (*Nagy, 1973, 1975, 1976*).

Egy további probléma – különösen a képességesztek esetében –, hogy a tanulók között nagyobbak a különbségek, mint amekkorát egy kötött formátumú teszttel le lehet képezni. Ha a teszt túl széles spektrumot próbál átfogni, minden tanuló csak néhány olyan feladatot talál, amelyik tudásszintjéhez közel áll, a feladatok többsége pedig vagy túl könnyű, vagy túl nehéz. Ha a tanulók a feladatokból egyénileg a képességszintjükhöz közeli válogatást kapnak, pontosabban be lehet határolni a konkrét fejlettséget.

A klasszikus tesztelmélet által kínált eljárásokat alkalmazva ki lehet számítani a teszt sokféle jellemzőjét, azonban a paraméterek többsége szigorúan véve csak a teszt bemérésére alkalmazott minta (tanulócsoporthoz) esetében lesz érvényes. A már korábban említett, és további, itt nem elemzett problémák megoldására a klasszikus tesztelmélet kereteit továbbfejlesztve illetve a PP tesztek kötött formátumát megbontva számos előreutató megoldás született. Azt a problémát azonban, hogy miként lehet feladatokhoz különböző paramétereket, mindenek előtt a nehézséget jellemző mértéket rendelni, függetlenül attól, hogy éppen melyik tesztben alkalmazzuk, a valószínűségi tesztelmélet (más neveken: valószínűségi tesztelmélet, modern tesztelmélet, Rasch modell, *Item Response Theory*, IRT) oldotta meg. Ezzel megnyílt az út a

változatos összetételű, kötetlen formátumú tesztek alkalmazása előtt. A valószínűségi tesztelmélet a mérés során elkövetett hibát és az itemek tulajdonságait más módon, nem determinisztikusan, hanem valószínűségi alapon kezeli.

A továbbiakban a klasszikus tesztelmélet képességszint-meghatározásának korlátain keresztül mutatjuk be a valószínűségi tesztelmélet modelljei közül kiemelve a Rasch modellel történő elemzések lehetőségeit, mely a technológiaalapú tesztelés adta lehetőségekkel ötvözve új távlatokat nyit a mérés-értékelés területén. E típusú elemzések jelentősége, hogy megvalósíthatóvá teszik a különböző, de horgony itemekkel ellátott tesztek és a teszteken mutatott teljesítmények összehasonlítását. Ha a kutatás során csak egyetlen egy tesztet alkalmazunk és az eredményeket nem viszonyítjuk más tesztek eredményeihez, nincs szükség ezen eljárások alkalmazására.

A klasszikus tesztelméleti elemzések során vagy a diákok nyerspontjai, vagy azok százalékos formában kifejezett értékei kerülnek összehasonlításra (mindkét esetben diszkrét pontok összevetéséről van szó). Két azonos képességet mérő teszt esetén azonban a tesztek nyerspont-értéke csak a diákok egymáshoz viszonyított sorrendjéről ad információt, de a közöttük lévő képességszintbeli távolságról nem, miután az annak függvényében változik, hogy könnyű vagy nehéz tesztet oldottak meg a diákok (Molnár, 2013b). Egy könnyebb és egy nehezebb, ugyanazon képességet mérő teszten elért összpontszámok közötti kapcsolat nem lineáris (Wu, 2006), két különböző nehézségű teszten nyújtott nyerspont-alapú teljesítmény direkt összehasonlítása nem releváns. A klasszikus tesztelmélet eszközszerével a diákok képességszintjének meghatározása jelentős mértékben függ a kutatás során alkalmazott teszt(ek) nehézségi szintjétől.

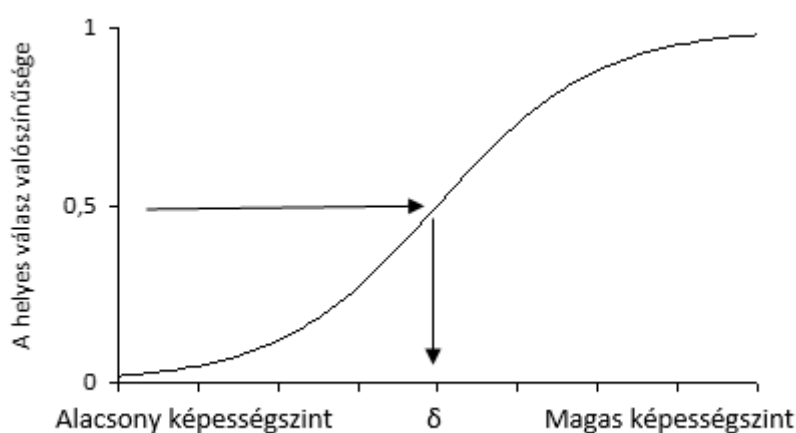
A nyers- vagy százalékpontok használatának további problémája a teszt itemei nehézségi szintjeinek és a diákok képességszintjeinek összekapcsolása. Egy ideális mérés során elvárjuk, hogy ha egy diák pl. 55 pontot ér el 100 pontból, akkor meg tudjuk mondani, hogy mit tud, az adott képesség fejlődésének milyen stádiumában van, mi várható el tőle. Ha nyers adatokat használunk a tanulók képességszintjének és az itemek nehézségi szintjének meghatározásakor, nem egyértelmű, hogy hogyan kapcsoljuk össze a két skálát (Molnár, 2013b).

A teszt, illetve itemek nehézségi szintjétől független képességszint-meghatározás előfeltétele egy olyan nyerspont-transzformáció, egy olyan matematikai függvény alkalmazása, ami megszünteti a teszt nehézségétől függő képességeloszlást. Erre alkalmas matematikai összefüggést biztosít a valószínűségi tesztmodellek közé sorolható Rasch modellben használt logisztikus függvény. A Rasch modell logisztikus transzformációja a nyers adatokat egy olyan skálára transzformálja, ami nemcsak a diákok közötti sorrendet, hanem a diákok közötti távolságok nagyságát is megőrzi.

A Rasch modell azon a feltételezésen alapul, hogy az adatokban kell lenni egyféle logikus hierarchiának (kevesebb mint / több mint): „a magasabb képességszintű személy nagyobb valószínűség mellett old meg bármely típusú itemet, mint a többi személy és hasonlóan egy item akkor nehezebb, mint a másik, ha bárki nagyobb valószínűséggel oldja meg a másik itemet, mint azt” (Rasch, 1960. 117. o.). Azokat az itemeket veszi nehéznek a modell, amelyeken kevesebben teljesítenek jól, és azokat sorolja a könnyűek közé, amelyeket sokan jól megoldanak. Az item nehézségét az adja meg, hogy milyen képességszint szükséges ahhoz, hogy  $p=0,5$  legyen a helyes megoldás valószínűsége, azaz az item nehézségét az azt 50% valószínűséggel megoldó egyén képességszintje határozza meg. Különböző képességszintek mellett az egy itemre adott helyes válaszok valószínűségét tipikusan az item karakterisztikus

görbéje írja le (2.1. ábra). A 2.1. ábra iteme esetében a ' $\delta$ ' átlagos képességszintű diák felel meg a korábban említett követelménynek, azaz ezen item nehézségi indexe ' $\delta$ '.

Miután az itemek nehézségi indexei a személyek képességszintjei alapján definiáltak, ezért az itemek nehézségét és az egyének képességszintjét közös képességskálán tudjuk ábrázolni. Ha ismerjük egy személy képességszintjét, meg tudjuk mondani, hogy milyen valószínűséggel oldana meg olyan itemet, amely nehézségi indexe értelmezhető a közös képességskálán, anélkül, hogy az adott személynek a valóságban meg kellene oldani azt az itemet. A  $\theta$  képességszintű tanulóhoz minden egyes item esetén hozzá lehet rendelni egy valószínűségi szintet, amilyen valószínűség mellett ő sikeresen oldja meg az adott itemet. Ennek következtében minden egyes személyhez hozzárendelhető annyi valószínűségi szint, ahány itemről van szó, illetve minden egyes itemhez hozzárendelhető annyi személyparaméter, ahányan a mintában vannak.



2.1. ábra  
Az itemkarakterisztikus görbe

Minden egyes valószínűségi tesztelméleti modell e valószínűségi értékeket használja fel a diákok elvárt teljesítményének és válaszmintázatának meghatározásakor, illetve minden egyes item minden tanulóhoz való hozzárendelése során is (Griffin, 1999). Ha minden egyes item esetén le tudjuk írni, hogy milyen képességszint szükséges 50%-os valószínűséggel történő megoldásához, akkor könnyen meg tudjuk határozni, hogy egy adott képességszintű diák milyen szinten van az adott képességterületen.

A valószínűségi tesztelméletben rejlő lehetőség, miszerint a mintában minden egyes diákról megmondható, hogy ő az adott, közös feladatbankban lévő feladatot milyen valószínűség mellett oldaná meg, még akkor is, ha a konkrét feladat megoldására nem kerül sor, túlmutat a klasszikus tesztelmélet határain. A klasszikus tesztelmélet eszközrendszerével kizárólag azon feladatok és tesztek eredményeiről, megoldottságáról beszélhetünk, és csak azokat elemezhetjük, azon teszteredményekből vonhatunk le következtetéseket, amelyeket a valóságban is megoldott a diák. Arról nem mondhatunk semmi, hogy ugyanazon diák, esetleg ugyanazon a vizsgált képességterületen egy könnyebb, vagy egy nehezebb teszten hogyan teljesített volna.

A Rasch modell a valószínűségi tesztelméleti modellek között csak egy modell, mégis speciális tulajdonságainak köszönhetően kiemelt szerepet kap a pedagógiai kutatások során (l. *Bond és Fox*, 2001, 2015; *Molnár*, 2005, 2006a, 2008a). A speciális objektivitás biztosítja, hogy a mintában bármely két személy összehasonlítása független attól, hogy az adott konstruktumot mérő itemek közül melyiken tesszük azt, illetve bármely e tulajdonsággal bíró két item összehasonlítása független attól, hogy milyen képességszintű személy oldotta meg azokat. Ez a féle objektivitás, függetlenség az IRT modellek közül csak a Rasch modell tulajdonsága, ami biztosítja a teszt és mintafüggetlen elemzéseket.

Az eredmények értelmezése során szem előtt kell tartani a Rasch modell néhány lényeges tulajdonságát. (1) A logit skála nem határozza meg a képességszintek és nehézségi indexek abszolút helyét, hanem a modell felállítja egyrészt a képességszinteken, másrészt a nehézségi indexeken, harmadrészt a képességszintek és nehézségi indexek között lévő relatív távolságokat. (2) A skálának nincsen abszolút nulla pontja. A két különböző skála egymáshoz való viszonyításának problémáját kiküszöbölhetjük, ha a két skálában van valami közös, összekötő elem (diák vagy item). Ebben az esetben lehetőség van a közös skála kialakítására, ahol már összehasonlíthatók a korábban külön skálázott itemek és személyek. (3) A logitegységnek nincs abszolút hossza. Ennek következtében az tesztfüggő, hogy az adott elemzés során milyen távol van egymástól két ember képességparamétere a képességskálán. Egy magasabb diszkrimináló erővel rendelkező teszt jobban széthúzza, jobban diszkriminálja a személyeket, mint egy, az adott mintát kevésbé diszkrimináló feladatlap. Ezért két egymástól független skálázás során nemcsak, hogy az egyes skála fokok, de a logitegység hossza sem hasonlítható össze, vehető azonosnak.

### 2.2.1. Horgonyzási technikák

A mérés során alkalmazott különböző tesztek átfedő, azonos, azaz horgony itemei, feladatai és a Rasch modell speciális objektivitás tulajdonsága lehetővé teszi a különböző teszteken mutatott teljesítmények összehasonlítását, másrészt a tesztfeladatok nehézségi fokának közös nehézségi skálán való kifejezését. Horgony itemek segítségével összehasonlítható például azonos minta által két különböző – más-más alkalommal megoldott – teszt eredménye, vagy különböző minták által azonos időpontban megoldott különböző tesztek eredményei. Mindkettő feltétele – két rész minta esetén –, hogy a horgony feladatokat minden egyes diák megoldja, több rész minta esetén, hogy a horgonyfeladatok kiosztása lehetővé tegye a feladatok összeskálázását. Az e módon létrejött, egymástól különböző tesztek nehézségei a horgony teszthez viszonyítva kerülnek meghatározásra, azaz összehasonlíthatóak lesznek.

A horgonyzásnak számos formája van: alkalmazhatunk horgony itemet, amelyet két teszt esetén mindkét teszt tartalmaz, vagy alkalmazhatunk horgony részteszteket, amelyeket két tesztnél maradvá mindkét teszt tartalmaz. Minél több közös item van a két tesztben, annál biztosabb a horgonyzás és minél kevesebb, annál lazább a kapcsolat. *Ridgway* (2003) horgonyzásra vonatkozó javaslata értelmében egy általános kognitív képességet mérő teszt mindkét alkalommal történő alkalmazása is lehetővé teszi a teljesítmények összeskálázását.

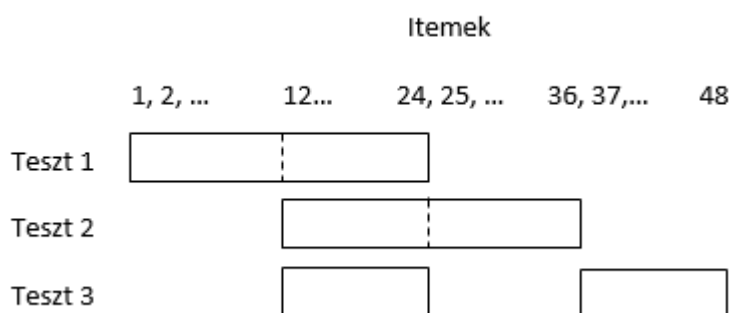
Módszertanilag a horgonyzás két alapvető módszerét különböztetjük meg: a szakértői döntéseken (Angoff eljárás, Jaeger módszer) alapulót és a statisztikai számításokon (pl.: Rasch skálázás) nyugvót. A horgonyzás legújabban használt módszerei ötvözik a két eljárást (pl.:

Ebel-módszer; Ridgway, 2003). Mindkét módszeren belül lehetőség van a mátrix design vagy a Ridgway nevéhez köthető kövér horgony (*fat anchor*) használata is.

A mátrix design egy általános, bármely típusú tesztelés során alkalmazható eljárás. Alkalmazhatóságát nem befolyásolja a teszt téje, vagy a tesztek száma. A továbbiakban a kutatásban kiköszvetítendő, azonos konstruktum mérésére alkalmas tesztek számának növekedése fényében mutatjuk be a horgonyzási lehetőségeket. Megfelelő erősségű horgonyzást érünk el azonos item számú tesztek esetén abban az esetben, ha mindkét tesztet két azonos hosszúságú résztesztre bontjuk és a résztesztek egyike azonos a két tesztben.

Ha a teszteket három résztesztre osztjuk, már gyengébb lesz a horgonyzás erőssége, azonban több itemet tudunk egymáshoz skálázni, miután csak az itemek harmadát kötik le a horgony itemek. A harmadolás módszerét abban az esetben érdemes használni, ha például nagyobb képességszintbeli különbség van a két különböző tesztet megírók között, mert akkor csak a teszt feladatainak harmada kell, hogy illeszkedjen a két minta képességszintjéhez, miután azok mindkét tesztben szerepelnek.

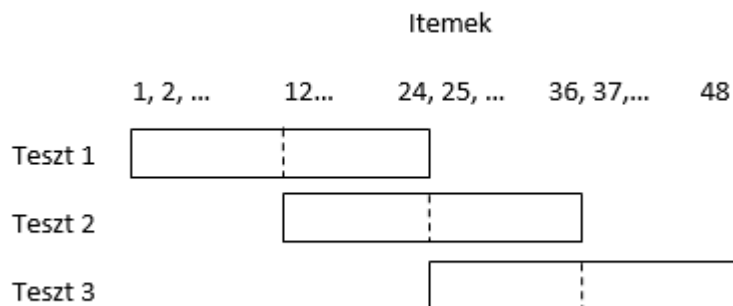
Három teszt esetén a teszteket két azonos itemszámú résztesztre bontva az összekapcsolás több módja adott: (1) Mindhárom tesztnek van közös résztesztje, azaz a vizsgálat során vannak olyan feladatok, itemek, amiket minden egyes diák megold. Ebben az esetben a tesztek többi iteme ezekhez az itemekhez skálázhatók (2.4. ábra). Ennél gazdaságosabb és tágabb életkori intervallumot átfogó mérések során is alkalmazhatóbb megoldás, amikor (2), a három teszt közül az egyik a másik két tesztet összekötő horgonyteszt (l. 2.2. ábra). Erre hazai alkalmazást l. Molnár (2003, 2006b, 2007) kutatásaiban.



2.2. ábra

*Három teszt horgonyzási lehetőségei (1 teszt 2 résztesztből áll – szűkebb képességtartomány mérése)*

Három teszt esetén (2.3. ábra), ha mindhárom tesztet három-három résztesztre osztjuk, alkalmazhatjuk az előbb említett két eljárás ötvöztetését. A horgonyzás során vannak olyan itemek, amelyeket mindhárom teszt tartalmaz, de bizonyos itemeket csak két-két teszt, illetve vannak olyan itemek, amelyek csak egy tesztet köthetőek. Ebben az esetben a tesztek résztesztjeinek eltérő egymáshoz rendelése adja a tesztelés rendszerének végső formáját. Gazdaságosság szempontjából a második lehetőség a kedvezőbb.

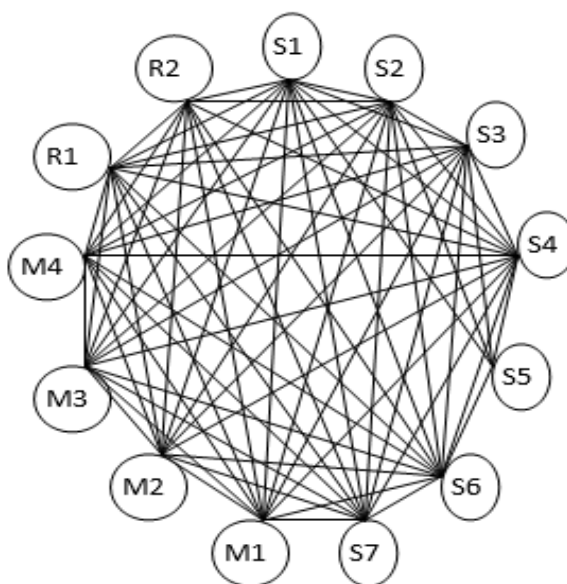


2.3. ábra

*Három teszt horgonyzási lehetőségei (1 teszt 2 résztesztből áll – tág képességtartomány mérése)*

Az OECD PISA 2006 vizsgálatban (OECD, 2009b) hasonló kutatási elrendezés mellett 13 különböző résztesztet használtak a tesztelés során, melyből az előző horgonyzási technikával 13 különböző tesztet állítottak össze a kutatók. Minden egyes teszt négy résztesztből állt, melyek különböző rotációját mutatja a 2.4. ábra. A kutatás szakértői külön figyelmet fordítottak arra, hogy minden egyes teszt a többi tesztel összeköttetésben legyen és minden egyes teszt a többi tesztel pontosan egy részteszt erejéig legyen horgonyzott.

A 2.4. ábra gráfként ábrázolva mutatja a PISA 2006 vizsgálat kutatási elrendezését. A gráf csúcsai a résztesztek, élei pedig az egyes résztesztek összekapcsolását, horgonyzását jelentik. A 13 kiközvetített tesztben előforduló összes kapcsolatot felrajzolva egy teljes gráfot kapunk, amelynek minden egyes csúcsát a többi csúccsal egy él köt össze. Minden egyes részteszt az összes többi résztesztel pontosan egyszer került horgonyzásra a kutatási elrendezés kialakítása során.



2.4. ábra

*A PISA 2006 mérés során alkalmazott klaszterkapcsolatok (S-természettudomány, M-matematika, R-olvasás; a számok a területen belül alkalmazott klaszter sorszáma)*

A rotáció másik lényeges tulajdonsága, hogy a résztesztek teljes teszten belüli elhelyezésével is törekedtek a kutatók a pozíciós hatás kizárására. Mind a 13 részteszt pontosan egyszer jelenik meg egy teljes teszten belül első, pontosan egyszer második, pontosan egyszer harmadik és pontosan egyszer negyedik pozícióban, azaz a teszt legelején, a teszt második negyedében, a teszt harmadik negyedében és végül a teszt végén. A rotációs mátrix egy további tulajdonsága, hogy páronkénti csoportosításban mindegyik részteszt pár csak egy tesztben fordul elő és többen nem.

#### 2.2.1.1. A kövér horgony

A horgonyzás másik lehetséges módja, amit nagy mintás, sztenderdekre épülő tesztelés esetén végezhetünk, a kövér horgony (*fat anchor*) használata. A kövér horgony egy olyan horgony itemekből álló halmaz, ahol minden egyes item a területet jól jellemző "ideális" item, valamint a fat anchor annyi itemet tartalmaz, amivel az egész, mérni kívánt terület lefedhető és jól mérhető. Az itemek és az azokból álló feladatok a vizsgált terület meghatározásán, fogalmi struktúráján, alapelvein, keretrendszerén nyugszanak. Ebből adódóan a kövér horgonyban lévő itemek mennyisége több, mint ami egy résztesztbe vagy tesztbe, egy tesztelési idő alatt tesztelt item mennyiségbe beleférne. A kövér horgony nem alkalmas egyedi teljesítmény mérésére és leírására, a vizsgálni kívánt populáció képességszintje jellemezhető vele. Nagyságából adódóan egyetlen egy diák sem oldja meg a kövér horgony összes itemét, feladatát, viszont a célpopuláció minden egyes diákjának meg kell oldani néhány feladatot, itemet belőle. Ezzel a módszerrel létrehozott adatbázis elemzése a sztenderdek és a teljes teszt konstruktum validitásának elemzésére alkalmas (*Ridgway, 2003*).

Több évfolyamot átfogó mérés esetén az összehasonlíthatóság és közel azonos nehézség biztosítása céljából az érintett összes évfolyamra jellemző, az adott évfolyam és terület sztenderdjeinek megfelelő feladatokat is tartalmaznia kell a kövér horgonynak. Ebből következőleg a horgonyzás során már két szempontot is figyelembe kell venni. Egyrészt a horgony itemek között legyenek a kövér horgonyból való itemek, másrészt legyenek olyan itemek, amik biztosítják a különböző évfolyamok közötti átmenetet, azok közös skálára hozását, összehasonlíthatóságát. Ebben az esetben az eredmények évfolyamtól függetlenül viszonyíthatóak az "ideális teszt"-hez.

### 2.3. A valószínűségi tesztelmélet és a magyar alkalmazások

Magyarországon az empirikus neveléstudományi kutatások a második világháború után indultak fejlődésnek. Az ötvenes évek végén, a hatvanas évek elején *Kiss Árpád* nevéhez köthető többek között az első jelentősebb hazai tudásszintmérés (*Kiss, 1960a, 1960b, 1960c, 1961*) és az Országos Pedagógiai Intézetben szerveződő mérésel-értékeléssel kapcsolatos kutatások elindítása. A hatvanas évek második felétől két pedagógiai értékeléssel foglalkozó központ bontakozott ki Magyarországon.

Az egyik a *Kiss Árpád* körül kialakult mérés-értékeléssel foglalkozó kutatócsoport (l. *Kiss, 1969, 1978; Báthory, 1973*). Ide köthetők az OPI (Országos Pedagógiai Intézet) által szervezett TOF országos reprezentatív felmérések (l. *Báthory, 1983*), illetve az OPI, majd OKI



(Országos Köznevelési Intézet), ill. KÁOKSZI (Kiss Árpád Országos Köznevelési Szolgáltató Intézmény) által szervezett Monitor-vizsgálatok (Vári, 1997).

A másik jelentős pedagógiai mérés-értékeléssel foglalkozó kutatóközpont kialakítása Nagy József nevéhez köthető. Vezetése alatt az akkori József Attila Tudományegyetem (ma Szegedi Tudományegyetem) Pedagógiai Tanszékén számos – egyrészt a tudás egyszerűbb elemeinek vizsgálatára (Nagy, 1971, 1973), másrészt különböző tantárgyak teljes tananyagát lefedő tudásszintmérő – teszt került kidolgozásra, illetve országos reprezentatív mintán bemérésre. A tesztek ismertető Standardizált témazáró tesztek sorozat 18 kötete 1973 és 1975 között jelent meg. A sorozat utolsó összefoglaló kötete a tesztfejlesztés módszereit tárgyalja (Nagy, 1975). Ezzel lerakta az objektív iskolai visszacsatoló mechanizmusok kiépítésének alapjait. Ugyanezen iskola keretein belül a 80-as években sor került több képesség táj életkori intervallumot átfogó fejlődésének vizsgálatára is (Nagy, 1987; Nagy és Gubán, 1987), amit az újszerű tudásszintmérő koncepciók kidolgozása és megvalósítása (l. diagnosztikus értékelés, Vidákovich, 1987, 1990), majd az alpműveltségi vizsgaközpont megalapítása és az alpműveltségi vizsgát előkészítő munkálatok, tesztek, feladatbankok kidolgozása (Nagy, 1997) követett.

Mindezen reprezentatív mintán alapuló kutatások klasszikus tesztelméleti eszközöket alkalmaztak az eredmények elemzése során, a valószínűségi tesztelmélet (*Item Response Theory*; IRT) eszközrendszerét, a benne rejlő lehetőségeket nem használták ki a kutatók.

A Kiss Árpád nevével fémjelzett vonulathoz kötődően a Monitor-vizsgálatok utódjaként és a PISA-vizsgálatok nyomán – azok módszereit, elemzési technikáit, alkalmazott eljárásait követve – 2001-ben elindultak az Országos kompetenciamérések (Balázs, Rábainé Szabó, Szabó és Szepesi, 2005; Balázs, Felvégi, Rábainé Szabó és Szepesi, 2006; Balázs, Ostorics és Szalay, 2007; Balázs, Ostorics, Szalay és Szepesi, 2010; Balázs és Ostorics, 2011; Balázs, Lak, Szabó és Vadász, 2013). A vizsgálatok tervezésénél, a tesztek összeállításánál, a képességszintek meghatározásánál, a közös képességskála kialakítása során már nélkülözhetetlen szerepet kapott a valószínűségi tesztelmélet, ami nélkül a kutatás mai felépítése, kivitelezése megvalósíthatatlan lenne. A nemzetközi kutatásokban használt elemzési technikák hazai integrációját segíti, hogy ugyanazon elemzők végzik a PISA vizsgálatok hazai vonatkozású elemzéseit, akik a Kompetenciamérések megvalósításáért felelősek.

A 90-es évek nemzetközi kutatásai egyre intenzívebben vetették fel a valószínűségi tesztelmélet használatának fontosságát és igényét. Az első magyar nyelven írt rövid ismertető Falus Iván szerkesztette (1993) könyv tudásszintmérő tesztekéről szóló fejezetében jelent meg (Csapó, 1993). Ezt követte Horváth György (1997) doktori értekezése alapján elkészített „A modern tesztmodellek alkalmazása” című könyve. A könyv néhány elvi és gyakorlati témát érintő problémát tárgyal.

A hazai szakmai folyóiratokban megjelent tanulmányokat tekintve a valószínűségi tesztelmélet használata, illetve szélesebb körű használatának igénye először Csapó Benő 1994-es, majd 2001-es tanulmányában fogalmazódik meg: „A kompetenciák széles életkori intervallumot átfogó vizsgálatához tehát szükség lenne olyan tesztrendszerek kifejlesztésére, amelyek egymástól kissé különböző nehézségű, egymásra épülő tesztek tartalmazzanak. A korábbi IEA vizsgálatok alapján az IRT eszközrendszerét felhasználva [...] már kifejlesztettek egy olyan eljárást, amely alkalmas volt arra, hogy a különböző korosztályok eredményeit egy skálán jelenítsék meg. A későbbiekben ilyen megoldások alkalmazásával kiterjeszthetnénk a képességvizsgálatok alkalmazási lehetőségeit is.” (Csapó, 2001. 389. o.) Ezt követően

különböző területeken találkozhatunk alkalmazásával [a teljesség igénye nélkül: problémamegoldó képesség skálázása (pl.: *Molnár*, 2003), induktív gondolkodás fejlődésének skálázása (pl.: *Molnár és Csapó*, 2011), idegennyelvi skálák építése (*Molnár, Nikolov és Csapó*, 2005), a nyelvi érettségi feladatainak skálázása (*Vígh*, 2008), nyelvi skálázás és feladatbank létrehozásának lehetősége (*Szabó*, 2008), olvasási képesség fejlődése (*Molnár és Józsa*, 2006)].

A 2000-ben indult PISA vizsgálatok eredményeinek elemzésekor a nemzetközi szakértőcsoportok a skálák megalkotása és a nemzeti eredmények ismertetése során már rutinszerűen használták a valószínűségi tesztelmélet eszközrendszerét. Ennek következtében a PISA vizsgálatok eredményeinek hazai ismertetése sem nélkülözhetette azt. Ez megteremtette annak a feltételrendszerét, hogy a hazai oktatáspolitikusok, pedagógusok és kutatók széles körben olvashassanak a közös skálák megalkotásának lehetőségéről PISA kontextusban (l. pl.: *Balázs, Ostorics és Szalay*, 2007).

A PISA elemzések során elsajátított eljárásokat, mérési és elemzési technikákat felhasználva hazánk 2001-ben elindította mára világszínvonalúvá vált nemzeti értékelési rendszerét a Kompetenciamérést. A több életkorban történő Kompetenciamérés elemzéseinek 2010-re már alapjává vált a valószínűségi tesztelmélet eszközrendszere. A mérési azonosító 2008-ban történő bevezetése lehetővé tette, hogy az egymást követő évek mérési eredményei tanulói szinten követhetők legyenek, bármelyik iskolába folytatja tanulmányait is a tanuló, azaz megvalósítható a diák longitudinális követése. Az ugyanazon diák által, de különböző években (évfolyamokon) megoldott tesztek eredményeinek összehasonlításához azonban még az IRT használatát megkívánó közös képességskálák kialakítása is szükséges volt.

Mindezzel párhuzamosan a *Nagy József* nevéhez kötődő Szegedi Iskola keretein belül, a munka folytatásaként a 90-es években a *Nagy József* által alapított, majd *Csapó Benő* által vezetett Képességfejlődés Kutatócsoport (KKCS) és a *Csapó Benő* által alapított Oktatásméleti Kutatócsoport longitudinális (*Csapó*, 2007), illetve tág életkori intervallumot átfogó kutatásai (l. pl.: *Csapó*, 2001; *Molnár és Csapó*, 2011) kikényszerítették a valószínűségi tesztelmélet mindennapos alkalmazását és használatát. Ennek következtében a két legjelentősebb hazai mérés-értékelési központban mára már rutinná vált a valószínűségi tesztelmélet eszközrendszerének használata. A két kutatóközösségen kívül kevés hazai alkalmazásával találkozunk.

## 2.4. A technológiaalapú tesztelés különböző szintjei

A technológia mérés-értékelésben betöltött szerepe sokrétű lehet, ebből adódóan nincs egységes definíció arra vonatkozólag, hogy mit jelent a technológiaalapú tesztelés. Attól függően, hogy a mérés-értékelés folyamatának melyik stádiumában, szintjén, vagy szintjein jelenik meg, továbbá a folyamatban résztvevők közül kik és mire használják az adott eszközt – aminek milyensége szintén definiáló erővel bírhat –, különböző definíciókat fogalmazhatunk meg.

A hagyományos papíralapú vizsgálatokban is már nélkülözhetetlen szerepet tölt be a számítógép. Általában azon történik a feladatok végső formába öntése, az adatok rögzítése, elemzése, a visszajelzések elkészítése stb. Ennek ellenére ezeket a méréseket a továbbiakban nem tekintjük technológiaalapúnak. Technológiaalapú mérés-értékelésről abban az esetben beszélünk, ha maga az adatfelvétel során is és nemcsak előtte, illetve utána kap fontos szerepet a technológia.

Az adatfelvétel tekintetében öt általános helyzetet különböztethetünk meg aszerint, hogy milyen területen, helyzetben milyen relevanciával alkalmazhatjuk a különböző eszközöket a vizsgálat során. Ezekben belül további esetek definiálhatók annak függvényében, hogy mi az a konkrét technológia, eszköz, amit alkalmazunk, és azt hogyan, milyen tesztelési módszerrel tesszük.

Az első esetben a mérésben résztvevő diákok olyan technológiai eszközöket használnak az adatfelvétel során, amiket más esetben, az iskolában, tanórán is alkalmaznak az adott területen kapott feladat, vagy probléma megoldása során. Például egy matematika teszt esetén természetes számukra a tanórán is gyakran használt számológép, vagy különböző grafikai ábrázoló programok, esetleg Excel táblázatok használata, még akkor is, ha a feladatmegoldás folyamata papíralapon történik. Ezek az eszközök kötődnek számukra a matematikai jellegű problémák megoldásához, ugyanakkor más helyzetekben általában nem használják ezeket. Ez esetben az adatok rögzítésében nem, de az adatfelvétel során szerepet játszik a technológia (Csapó, Ainley, Bennett, Latour és Law, 2012).

A második esethez olyan területek vizsgálata, olyan adatfelvételi helyzetek tartoznak, ahol alapvetően nem jellemző, vagy nagyon ritka, hogy a diákok egy technológiai eszközt használnak. Ilyen például az esszéírás, amit az iskolában a diákok általában papíralapon végeznek. Ebben az esetben, ha egy diák nem használ számítógépet, kezelése, az azon folytatott munka – a konkrét példánál maradva a mondat szerkesztés és gépelés – számára nehézkes lehet, ami negatívan befolyásolhatja eredményét. Aki többet ír billentyűzeten, mint papíron, esetleg, aki blogot vezet, előnyben van, nemcsak a gépelés gyorsasága, hanem a mondatformálás terén is. Ebben az esetben a papíron és számítógépen írt esszé nem összehasonlítható (Powers és Potenza, 1996). Az adatfelvétel tervezése során háromféle módon járhatunk el:

- minden egyes diákot hagyományos papír-ceruza módon tesztelünk, hogy lássuk ott milyen hatékonyak;
- minden egyes diákot technológiaalapon tesztelünk, hogy lássuk az adott területen milyen hatékonyak az adott eszköz használatában;
- minden diákot azon a módon tesztelünk, ahol rutinosabb, amit megszokott (Horkay, Bennett, Allen, Kaplan és Yan, 2006 idézi Csapó, Ainley, Bennett, Latour és Law, 2012).

A harmadik adatfelvételi szituáció azokra a területekre jellemző, ahol nélkülözhetetlen az adott technológiai eszköz használata, anélkül értelmetlenné válna a terület mérése. Például számítógép nélkül nem lehet hatékonyan tanítani a programozást. Ezekben az esetekben a mérés-értékelés során is kézenfekvő az adott eszköz használata (Bennett, Persky, Weiss és Jenkins, 2007; Bennett, Jenkins, Persky és Weiss, 2003).

A negyedik típusú kutatási helyzet célja lehet annak számszerűsítése, hogy a diákok képesek-e az adott feladat, probléma megoldásában jobb teljesítményt elérni, ha egy technológiai eszközt a rendelkezésükre bocsátunk, mint anélkül. Ha egy új, a diákok számára kevésbé ismert és használt eszközt adunk részükre, akkor opcióként adhatjuk azt a lehetőséget is, hogy nem veszik igénybe az adott eszközt és anélkül oldják meg a feladatot (l. pl.: Molnár, 2008b). Az e témakörbe sorolható kutatások a technológiai eszközök tanulást és tanítást befolyásoló szerepét vizsgálják (l. pl.: minden diák és tanár számára biztosítanak 1-1 laptopot, l. Bebell és Kay, 2010; Suhr, Hernandez, Grimes és Warschauer, 2010; Bebell és O'Dwyer, 2010).

Az ötödik esetben a technológia a diákok közötti együttműködést, kollaborációt és a hatékonyabb tudásépítést és tudásmegosztást támogatja, miután a 21. század társadalmában

mind a tudásépítés, mind a tudásmegosztás nélkülözhetetlen eszköze a technológia, a közösségi szinten történő aktív kommunikáció (*Scardamalia, 2002; Lipponen, 2002*). A számítógéppel támogatott kollaboratív tanulás, azaz az együttműködő tanulás információs és kommunikációs technológia segítségével történő támogatása számos hazai és nemzetközi kutatást indukált (pl.: *Molnár P., 2009; Dorner, 2007; Molnár és Kárpáti, 2009; Neo és Neo, 2001; O'Malley, 1995; Macdonald, 2003; Kyllonen, 2009*).

#### 2.4.1. A technológia szerepe a mérés-értékelés hatékonyságának növelésében

A mérés-értékelés egész folyamatában – a feladatírástól az eredmények visszajelzéséig – szerepet játszhat a technológia. Három esetet különböztetünk meg aszerint, hogy a technológia alkalmazása befolyásolja-e a tesztelés validitását, azaz ugyanazt mérjük-e papír-, mint technológiaalapon.

Ideális esetben (tévesztés, elírás, elütés faktorát kizárva) a tesztelés eredményét és tárgyát, a tesztelés validitását nem befolyásolja a technológia, ha azt az adatfelvétel kivételével bármely más esetben alkalmazzuk. Például

- a) feladatírás, tesztfejlesztés során,
- b) automatikus, vagy félautomatikus itemgenerálásra,
- c) a feladatok megosztására, bírálatára, javítására (*Bejar, Lawless, Morley, Wagner, Bennett és Revuelta, 2003*), vagy
- d) az eredmények rögzítésére, kódolására, elemzésére,
- e) a visszajelzések elkészítésére és visszajuttatására, amivel nyomtatási és szállítási költséget is megtakarítunk.

A második esetben kevés információt, kevés olvasnivalót tartalmazó papíralapú feladatok az eredetivel megegyező formában való digitalizálása történik. Ebben az esetben csak a feladatokat közvetítő eszköz, vagyis a médium változik meg. A feladat a papír helyett a képernyőn jelenik meg, a válaszadás billentyűvel, egérrel, érintőképernyővel, vagy valami egyéb elektronikus eszközzel történik. A tesztelés továbbra is lineáris marad, a feladatok azonos sorrendben jelennek meg minden egyes tesztelt személy előtt. Érintőképernyőt használva a papíralapú teszteléshez egészen közeli hasonlóságot lehet elérni, a vizsgázó – az érintőképernyő technológiájának függvényében – egy digitalizáló vagy egy közönséges toll segítségével jelöli meg választát. Egér vagy billentyűzet használata esetében már szükség van némi technikai készségre. A digitalizálásra kerülő feladatok köréből ebben az esetben kizárjuk a sok információt, sok olvasnivalót, ezért esetlegesen görgetést, vagy hosszabb szövegek beírását, begépelését kívánó, ezért a papíralapú feladatokhoz képest más típusú információfeldolgozást igénylő feladatokat. A legtöbb létező számítógép-alapú teszt ehhez hasonló formátumú, feleletválasztós feladatokból álló standardizált teszt (*Jurecka és Hartig, 2007*). A számítógépes tesztelés e szintje is számos előnnyel jár (l. 2.5. részt). A technológiát használhatjuk például

- a) a tesztek diákokhoz való eljuttatására, amivel jelentős nyomtatási, tárolási és szállítási költséget takarítunk meg. Ebbe az esetben sorolható még, ha
- b) digitális audio vagy videó fájlokat osztunk meg a diákokkal – gyakori tesztelési mód idegen nyelvi mérések során, amikor a diákok szövegértését mérik –, amelyeket korábban magnón vagy videón keresztül hallgattak rosszabb minőségben a diákok.
- c) Hasonlóképpen ide sorolható a diákok beszédképességének mérése, amelyet korábban magnóval, vagy videomagnóval rögzítettek (*Bennett, Goodman, Hessinger, Ligget,*

*Marshall, Kahn és Zack, 1999*). Ennek költsége a fent nevezett okok miatt szintén magasabb, mintha a számítógép segítségével digitálisan rögzítjük a diákok beszédét. Mindkét esetben a jobb minőség alacsonyabb előállítási költséggel párosul. Előbbi esetben nem kell kazettákra másolni a hanganyagot, elegendő azt online eljuttatni az iskolákhoz, utóbbi esetben akár párhuzamosan is történhet az adatfelvétel és a teszteléshez nem szükséges külön, plusz eszköz fizikai szállítása.

- d) Megfelelő tesztelési módszer (pl.: feleletalkotó itemek, időmérés) alkalmazása mellett még az sem szükséges, hogy a tesztelt személyek fizikailag egy helyen legyenek, amivel szintén jelentős költségmegtakarítást érhetünk el (*Zhang, Powers, Wright és Morgan, 2003*).
- e) A számítógépes tesztelés e szintjén lehetőség nyílik automatikus itemgenerálásra is, így bizonyos típusfeladatok mindig új formában jelenhetnek meg (például a szöveges feladatokban mindig más-más számértékek, nevek szerepelnek; *Gierl és Haladyna, 2013*).
- f) A technológia lehetőséget ad a válaszok azonnali és automatikus kiértékelésére, amivel jelentős emberi munkát takarítunk meg (*Williamson, Mislevy és Bejar, 2006*), illetve felgyorsítjuk a visszajelzés folyamatát, növelhetjük a tesztelés objektivitását, ezáltal validitását, ami szorosan összefügg a tesztelés hatékonyságával (*Csapó, Molnár és Nagy, 2014*).

A harmadik esetben a mért konstruktum már különbözik attól, amit papír-ceruza alapon mérhetünk, mérnénk, azaz a technológia adatfelvétel során történő alkalmazása jelentősen befolyásolja azt, amit mérünk. Ebben az esetben alkalmazhatunk olyan

- a) interaktív szimulációkat, kísérleteket, amiket technológiai eszköz nélkül nem tudnának elvégezni a diákok, majd kérhetjük azok leírását, vagy azokkal kapcsolatos feladatok, problémák megoldását (*Bennett, Persky, Weiss és Jenkins, 2007*).
- b) Ide sorolható az a féle tesztelés is, amikor a diákok tanulnak a tesztelés során. A diákok feladatokra, problémákra adott válaszaik meghatározzák a tesztelés következő problémáját, azaz a tesztelés során dinamikus a feladatadás (l. 2.4.3. részt). A visszajelzésből tanulnak a diákok.
- c) E csoportba sorolhatóak azon papíralapú tesztekhez hasonló tesztek, mely feladatai sok információt, sok olvasnivalót, sok képet, esetlegesen sok írni/ gépelni, vagy számolni valót tartalmaznak, azaz olyan feladatelemeket, olyan típusú, tulajdonságú feladatokat, amelyek médiahatást idéznek elő (l. *Hülber és Molnár, 2013*; *Csapó, Molnár és R. Tóth, 2009*; *Leeson, 2006*).

#### 2.4.2. Az alkalmazott technológia típusának és módjának függvényében a technológiaalapú tesztelés típusai

A technológiaalapú tesztelés megvalósításának számos lehetősége ismert. A közvetítő eszköz és a kiközvetítés módjának kiválasztását számos tényező befolyásolja. A döntés során lényeges szempont, hogy a választott technológia biztosítsa az elemzéshez megfelelő minőségű adatok begyűjtésének lehetőségét.

A technológiaalapú tesztelés (*Technology Based Assessment – TBA*), illetve elektronikus tesztelés (*e-Testing*) magában foglalja az összes olyan mérési-értékelési rendszer alkalmazását, ahol az adatgyűjtésre valamilyen információs-kommunikációs technológiai eszközt használunk. Annak ellenére, hogy ez az eszköz általában a számítógép, mégis a számítógépes mérés-értékelés halmazát magába foglaló bővebb halmazként megkülönböztetjük ezt a

kategóriát. Ennek oka, hogy bizonyos esetekben a közvetítő eszköz nem feltétlen a számítógép, lehet mobiltelefon, szavazórendszer stb., amelyek egy része alkalmas arra, hogy a nap bármely időszakában bizonyos kérdéseket tegyen fel a mérésben résztvevőnek – attól függetlenül, hogy az illető helyileg hol van –, aki arra azonnal válaszolni tud.

A technológiaalapú mérésen belül – elterjedtségéből fakadóan is – a legtöbb lehetőséget a számítógép-alapú értékelés rejti magában, ahol az adatfelvétel közvetítő eszköze a számítógép, a mérőeszköz (teszt, kérdőív) kérdései, feladatai, itemei a számítógép monitorján jelennek meg. A válaszok bevitele is a számítógépen keresztül történik, a billentyűzet, egér, érintőképernyő stb. segítségével, ahol azok rögtön elektronikusan rögzítésre kerülnek, majd a válaszok elemzése is általában a számítógép felhasználásával történik. A számítógép-alapú tesztelésbe beletartozik annak mind hálózati, mind interneten keresztül történő alkalmazása. Ha semmilyen hálózatot (helyi hálózat, internet) nem vonunk be a tesztelés lefolytatásába, akkor a tesztelést végző programot, feladatlapot minden egyes számítógépre installálni kell. Az esetleges változtatásokat minden egyes számítógépen külön regisztrálni kell, majd az adatokat minden egyes számítógépről be kell gyűjteni. Ez sok tekintetben korlátozhatja a tesztelést. Ma már elérhető megoldás, hogy pendrájvról fusson a tesztelést végző program és ugyanoda mentődjenek a válaszadatok is. Ebben az esetben nincs szükség arra, hogy a helyi gépekre programokat installáljanak, elegendő a megfelelő USB kimenet és a program futásához szükséges környezet biztosítása. A tesztelés e típusa közép és hosszú távon olcsóbb, mint a papíralapú tesztelés lebonyolítása, sőt a pendrájvok később újra felhasználhatóak (*Csapó, Ainley, Bennett, Latour és Law, 2012*).

A hálózatalapú mérés egy gyakori alkalmazása, amikor az adott hálózaton belül egyszerre több gépen zajlik a tesztelés, a tesztelést egy külön számítógépről irányítják, ahol az adatok összegyűjtése, elemzése történik (*Csapó, Molnár és R. Tóth, 2008*). Ezt a külön, jelen esetben helyi szerverként működő gépet, notebookot, melyre a teszteléshez szükséges összes programot előzetesen installáltak, elegendő eljuttatni az iskolába, majd az iskola erőforrását, számítógépeit használva, erre csatlakozva történik a tesztelés. Ebben az esetben is két módja van a tesztelésnek. Vagy a tesztelés előtt minden egyes adatfelvételben részt vevő gépre felinstallálják a teszteléshez szükséges szoftvert, vagy a gépek a tesztelés alatt beállított szerverhez – hasonlóan az internet alapú teszteléshez, l. később – csatlakozva egy böngésző segítségével érik el a szükséges programokat. Az adatok begyűjtése történhet manuálisan (pl.: CD-re írva) vagy közvetlenül elektronikusan (pl.: ftp-t alkalmazva; *Csapó, Ainley, Bennett, Latour és Law, 2012*). A kiértékelés szoftvertől függően vagy a helyi számítógépen, vagy a központi szerveren történik.

Az internet alapú adatfelvétel irányítását egy központi szerver végzi, azaz a helyi számítógépekre nem kell felinstallálni a tesztelést végző programot. Mivel a szoftver külső gépen fut a gépekkel szemben felállított követelmények alacsonyabbak (*Csapó, Molnár és R. Tóth, 2008; Csapó, Ainley, Bennett, Latour és Law, 2012*), ugyanakkor megfelelő számú és minőségű internetkapcsolattal (és internetes böngészővel) rendelkező számítógépre van szükség. Az iskolában szükséges sávszélesség függ a tesztfeladatok jellegétől és az egyszerre tesztelt diákok számától. Ha az iskolában ez nem áll rendelkezésre, a probléma megoldható vezeték nélküli (wireless) internetkapcsolat alkalmazásával is, azonban ez jelentősen megdrágítja a vizsgálatot. Egy másik megoldás, ha proxy szerver programot alkalmazunk a tesztelés során. A proxy szerver program csak egyszer tölti le a központi szerverről a tesztelésben kiosztott feladatokat (nincs szükség arra, hogy az iskola minden egyes számítógépe

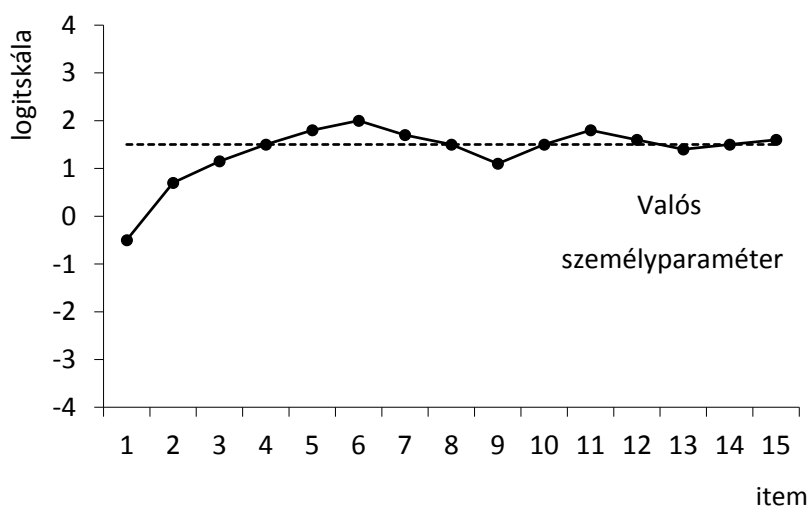
azt külön megtegye), majd azt a belső hálózat segítségével juttatja el a kliens számítógépekhez. Ennek előnye, hogy a kliens számítógépeken biztosított lesz a megfelelő sebességű tesztelés. Bármely megoldás esetén nélkülözhetetlen a megfelelő központi server alkalmazása, ami párhuzamos elérések esetén is biztosítja a tesztelés zavartalan lefolyását (*Molnár, Papp, Makay és Ancsin, 2015*).

Az internet alapú tesztelés legkifinomultabb, legtöbb információt szolgáltató formája az adaptív tesztelés. Az adaptív tesztelési technika személyre szabott, a diákok képességszintjéhez igazodó, kötetlen formátumú tesztek kiközvetítését valósítja meg.

### 2.4.3. Adaptív tesztelés

Az adaptív tesztelés a tesztelés korábbi, fix formátumú megvalósításához képest a teljesítmények sokkal finomabb felbontását, mérését teszi lehetővé. Adaptív tesztelés során a tesztet megoldó személy nem előre meghatározott feladatokat, előre meghatározott sorrendben old meg, hanem attól függően kapja a teszt egyes feladatait, hogy az előzőkön miképp teljesített. Ennek következtében az adaptív tesztelés hatékony megvalósítása internet alapú adatfelvételt igényel.

Feladat szintű adaptív teszt esetén, ha a tesztelt személy helyesen oldja meg a teszt egyik feladatát, következő feladatként egy nehezebbet kap, ha elrontja azt, akkor a teszt következő feladata számára egy könnyebb feladat lesz. Ezen eljárás következtében a tesztelés során személyre szabott tesztek dinamikus összeállítása valósul meg (*Molnár, 2013b*). Ezen algoritmus hatására adaptív tesztelés során a tesztelt személyek többségében olyan feladatokat kapnak, amelyek a lehető legtöbb információt szolgáltatják képességszintükről, miután ezek a feladatok a lehető legközelebb vannak a valós képességszintjükhöz (2.5. ábra).



2.5. ábra

*A feladatszintű adaptív tesztelés menetének illusztrációja. A pontok a feladatok nehézségi szintjét reprezentálják*

Adaptív tesztelés során a fix teszteknel tapasztalt mérési pontosság eléréséhez kevesebb feladatra, illetve rövidebb időre van szükség. Ugyanolyan mennyiségű feladat és idő

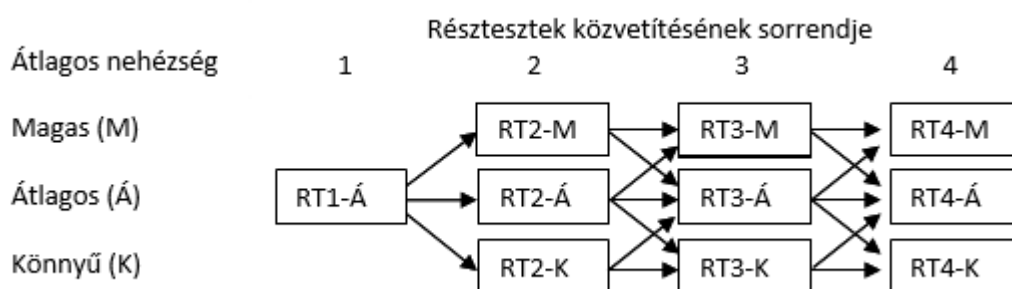
alkalmazása mellett viszont biztosított a nagyobb mérési pontosság (Molnár, 2013b; Frey, 2007). Az újabb és újabb itemek kiválasztása az előre meghatározott adaptív algoritmus szabályrendszere függvényében addig tart, amíg (1) az előre meghatározott mennyiségű itemek megoldásra kerültek; (2) a személyparaméter becslési hibája a megengedett hibahatáron belül mozog; (3) eltelt a tesztelésre fordítható idő; (4) az itembankban előforduló összes item bemutatásra került (Csapó, Molnár és R. Tóth, 2008).

Mindennek megvalósításában a Rasch modell (l. 2.2. részt) a feladatok, résztesztek paraméterezésében játszik szerepet. Adaptív tesztelés esetén egy korábban paraméterezett, indexelt feladatbank áll a tesztelés háttérében. A feladatokat korábban említett horgonyzási technikák segítségével tesztekbe kell sorolni, hogy az előzetes adatfelvétel lehetővé tegye az itemek empirikus mutatóinak (pl. nehézségi index, diszkriminációs index) meghatározását. Ha megtörtént az azonos konstruktumot mérő feladatok paraméterezése és azokból összeállítottuk a feladatbankot, indulhat az adaptív tesztelés megvalósítása. Miután a tesztelés eredménye közös nehézségi skálán definiált itemekből összeállított feladatbankon alapul, az eredmények viszonyíthatóak egymáshoz. Ha a tanuló részt vett már korábbi tesztelésben, ahol a teszt az adott feladatbank feladataiból került összeállításra, akkor a korábbi teljesítménye összevethető aktuális eredményével, még akkor is, ha összességében minden egyes alkalommal más itemeket oldott meg. A tanuló eredménye összevethető a többi diák azonos mérésben megoldott teszteredményével és a feladatbank felépítésének alapját képező tudományosan kidolgozott standardokkal is. A számítógépes adaptív tesztelés megvalósítását összességében a valószínűségi tesztelmélet, illetve speciális objektivitás tulajdonsága miatt a Rasch modell tette lehetővé. Adaptív tesztelés alkalmazásával kevesebb feladat, item használatával rövidebb idő alatt pontosabban meghatározható a tesztelt személyek képességszintje (Molnár, 2013b).

Ugyanakkor a feladat szintű adaptív tesztelés egyik fő problémája, hogy az itemek paraméterei változnak annak függvényében, hogy az adott feladat a teszt melyik részén helyezkedik el (elején, közepén, végén), milyen itemek veszik körül. Ez fix tesztelés esetén is jelentkező probléma – ugyanaz a feladat másként viselkedik attól függően, hogy a teszt melyik részén található. Ezen probléma megoldását kínálja a részteszt szintű adaptív tesztelés megvalósítása, ahol előre meghatározott, fix résztesztekkel történik ugyanezen algoritmus megvalósítása. A részteszt szintű adaptív tesztelés (*multistage testing*; Luecht és Nungester, 1998) során az itemszintű adaptív tesztelés kapcsán látott eljárás előre meghatározott, különböző nehézségű résztesztekkel valósul meg. A 2.6. ábra szemlélteti ennek egyik megvalósítási lehetőségét. A tesztelés előtt tíz különböző, de horgony itemekkel összekapcsolt és összességében három (könnyű, átlagos és nehéz) különböző nehézségi szintű résztesztet definiálunk. Mindegyik részteszthez hozzárendeljük, hogy a tesztelés melyik fázisában kerülhet közvetítésre (első, második, harmadik stb. lépés során). Az első lépésben minden diák ugyanazt a résztesztet kapja meg, majd annak függvényében, hogyan teljesített azokon a feladatokon második lépésként nehezebb, azonos nehézségű, vagy könnyebb résztesztet kap. Ez az eljárás folytatódik a tesztelés végéig, azaz jelen példa esetében, mindenkinek maximum négy részteszt erejéig.

A részteszt szintű adaptív tesztelés megvalósíthatóságának háttérében is a Rasch modell áll. Ha megtörtént az azonos konstruktumot mérő feladatok paraméterezése, a paraméterek alapján megvalósult a résztesztek összeállítása, létrejött a résztesztekbe sorolt feladatokat tartalmazó feladatbank, megtörtént az adaptív algoritmus kiválasztása, indulhat a részteszt szintű adaptív tesztelés megvalósítása.

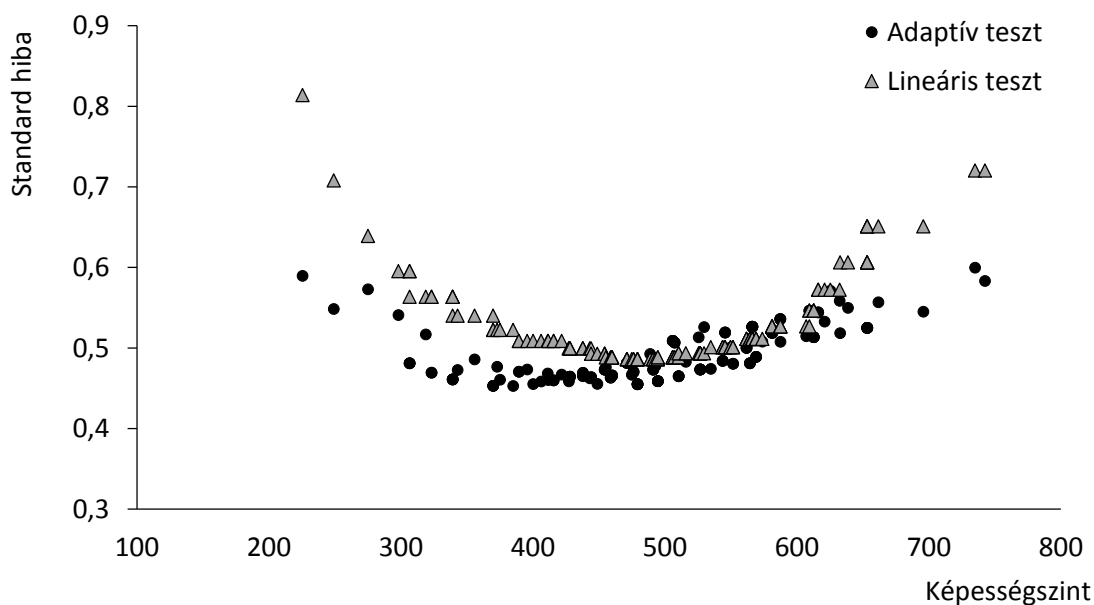




2.6. ábra

*Példa részteszt szintű adaptív tesztelés megvalósítására (RT: részteszt)*

Az adaptív tesztelés hasonló elven működik, mint a szóbeli felelés, amikor a pedagógus az előző válaszok függvényében teszi fel az újabb, könnyebb vagy nehezebb kérdéseket. Az adaptív algoritmus a megelőző itemeken vagy részteszteken nyújtott teljesítmény alapján választja ki a paraméterezett feladatbankból az újabb feladatot vagy résztesztet, az adaptivitás típusától függően. Szintén algoritmikusan meghatározott az is, hogy a tesztelés mely ponton, milyen teljesítményt mutatva ér véget (Molnár, 2010a; Csapó, Lőrincz és Molnár, 2012; Magyar és Molnár, 2013, 2014). Következésképpen az adaptív tesztelés a hagyományos, rögzített teszteknel sokkal pontosabb képességbecslésre képes, azaz csökken a mérési hiba (Magyar és Molnár, 2015; 2.7. ábra).



2.7. ábra

*Az adaptív és a lineáris formátumú teszt standard hibáinak alakulása a tanulók képességszintjének függvényében (Forrás: Magyar, 2014b. 59. o.)*

A mérési hiba csökkenésének mértéke függ a tesztet megoldó diák képességszintjétől. Átlagosan az alacsonyabb és magasabb képességszinttartományban nagyobb mértékű becslési

pontosság figyelhető meg, míg az átlagos képességszintű diákok esetében nem mutatható ki jelentős mértékű különbség a lineáris és adaptív teszteken elért eredmények között. Ennek oka, hogy a lineáris tesztek feladatai általában az átlagos képességszintű diákoknak megfelelő nehézségű szintű feladatokból épülnek fel. Mindemellett, mivel az azonos nehézségű itemek generálása random módon történik az akár több ezer itemet tartalmazó itembankból, minden tesztalany eltérő tesztverziót tölt ki, így lecsökken a lehetősége a tanulók összedolgozásának, a másolásnak, sűgásnak a tesztelés idején (Molnár, 2013b).

#### 2.4.3.1. Nő a tesztelés során kinyert információ mennyisége

A hagyományos, fix, kötött, valamint a kötetlen formátumú, a valószínűségi tesztelmélet adta eszközrendszer kihasználó adaptív tesztelés között jelentős mértékű különbség realizálódhat a tesztelés során kinyert információ mennyisége kapcsán, azaz a mérés pontossága tekintetében. Egy adott teszt, feladat, item kapcsán annál nagyobb a kinyert információ mennyisége, minél közelebb áll a teszt, feladat, item nehézségi szintje az egyén képességszintjéhez. Ugyanaz a teszt az egyik mintán nagyobb pontossággal mér, mint a másikon, ugyanarról a mintáról két különböző nehézségű teszt különböző mennyiségű és pontosságú információt szolgáltat (Molnár, 2013b).

A képességszint és a nehézségi index függvényében meghatározható a helyes válasz valószínűsége és az adatfelvétel során kinyert relatív információ nagyságát. Ha a személy képességparamétere alacsonyabb, mint az item nehézségi indexe, akkor a képességparaméter és a nehézségi index különbsége pozitív és a helyes válasz valószínűsége nagyobb, mint 50%. Minél nagyobb ez a különbség, annál közelebb van a helyes válasz valószínűségének nagysága az 1-hez, azaz a 100%-hoz (azt sohasem éri el). Ha az item túl nehéz az adott személy számára, azaz a képességparaméter és az itemnehézség különbsége negatív szám, akkor a sikeres megoldás valószínűsége kevesebb mint 50%. Abszolút értékben minél nagyobb ez a különbség, annál közelebb lesz a helyes válasz valószínűsége 0-hoz.

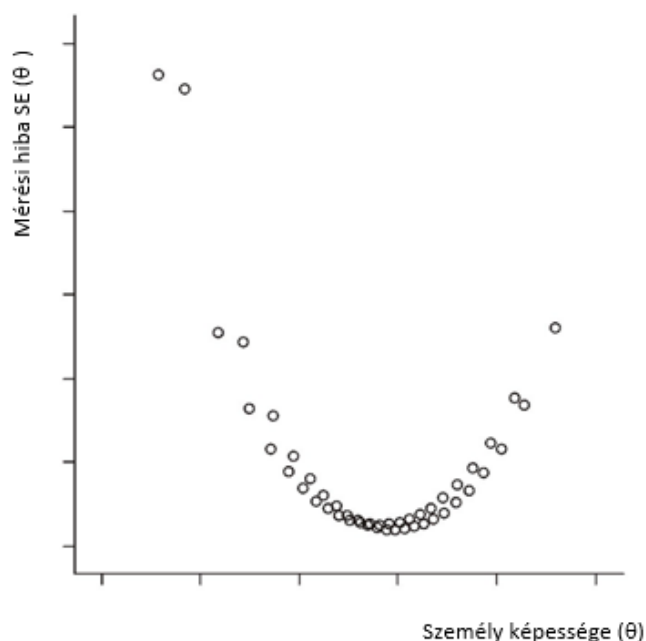
A kinyert relatív információt akkor tekintjük 100%-nak, ha a személy képességparamétere és az item nehézségi indexe megegyezik, azaz  $\theta_i - \delta_j = 0$  (Write és Masters, 1982). Ez a mérőszám annak eldöntését segíti, hogy az adott mérésben az adott item milyen mértékben járul hozzá a személy képességparaméterének meghatározásához, azaz mennyi információt veszítenénk az adott személy képességparaméterének meghatározása során, ha az adott itemet elhagynánk a tesztből. Ha az item nehézsége ( $\delta_j$ ) a személy képességparaméterének ( $\theta_i$ ) egy logitegységes környezetében van, akkor a  $\delta_j$  és a  $\theta_i$  –ről nyert információ több, mint 79%, ami fokozatosan 45%-ra csökken, ahogy a két paraméter közötti távolság nagysága 2 logitegységre nő. [Egy teszt információs függvényének meghatározásakor az egyes itemek információs függvényei összeadódnak ( $I_t = \sum_{i,j=1}^n I_{ij}$ , ahol  $I_t$  a teszt információs függvénye) (Adema és Gademann, 1992)].

Minél nagyobb a személy képességparamétere és az item nehézségi indexe közötti különbség nagysága, annál több item szükséges a személy képességparaméterének ugyanazon pontosságú meghatározásához. Ez az oka annak, hogy a legpontosabb információ kinyerését minden esetben azok az itemek biztosítják, amelyek nehézségi szintje közel áll a tesztelt személy képességszintjéhez.

Összességében megállapítható, hogy minél kevesebb item található a személy képességszintjének megfelelő nehézségi szinten, annál nagyobb mérési hibával dolgozunk,

annál pontatlanabb lesz a tesztelés során kinyert információ, annál kevésbé van összhangban a személy teszten mutatott teljesítménye és valódi képességszintje.

A mérési hiba, azaz a mérés pontosságának és a személy képességszintjének kapcsolatát adott nehézségi szintű teszt mellett a 2.8. ábra mutatja. Az ábra alapján megállapítható, hogy a mérés pontossága és a személy képességszintje közötti összefüggés nem lineáris. Ez az oka annak, hogy világszerte jelentős kutatások indultak a kötetlen, a személy képességszintjéhez illeszkedő, úgynevezett adaptív tesztelés megvalósítása érdekében, fokozatosan felváltva a hagyományos rögzített tesztelési technikákat.

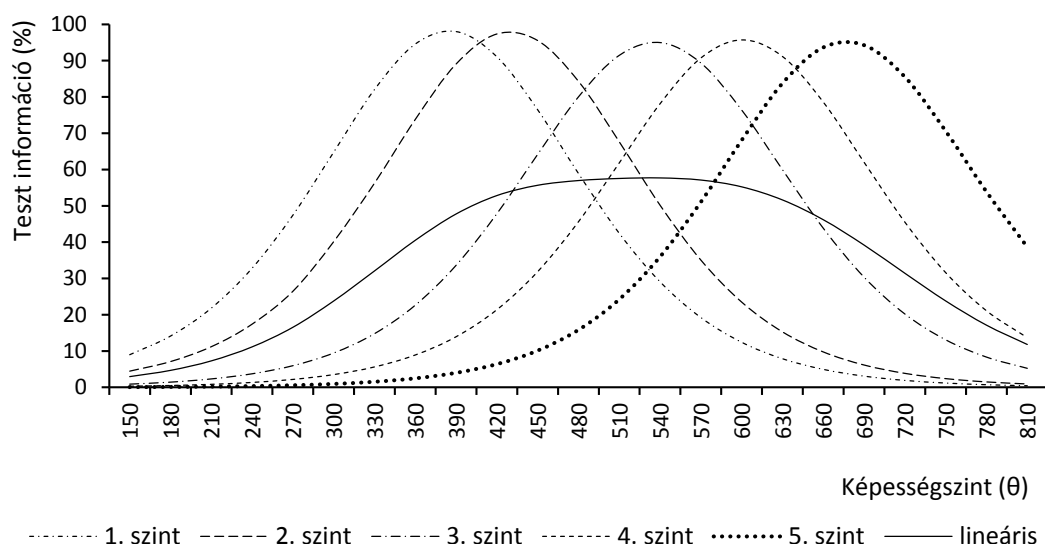


2.8. ábra

*A személy képességszintje és a mérési pontosság közötti összefüggés adott nehézségű teszt esetén (Jude, 2006 alapján)*

A valószínűségi tesztelmélet eszközszerét kihasználó, kötetlen, úgynevezett adaptív teszteléssel megvalósítható, hogy a teszt nehézsége összességében megfeleljen a személy képességszintjének, azaz a legtöbb és legpontosabb információt nyerjük ki a tesztelés folyamán. A 2.9. ábra ugyanazon feladatokból álló tesztrendszer lineáris és adaptív formában történő kiközzvetítésének hatékonyságát hasonlítja össze az adatfelvétel során kinyert információ mennyisége szempontjából.

Adaptív tesztelési technika alkalmazásával minden egyes képességszinten jelentős mértékben megnőtt a kinyert információ mennyisége, azaz pontosabbá vált a képességszintbecslés, jelentős mértékben csökkent a mérési hiba nagysága. Mindennek kivitelezése feltétele az azonnali visszacsatolás, a feladatok megoldásának azonnali értékelése, amely a hagyományos papíralapú technikákkal már nem kivitelezhető, megvalósításához technológiaalapú adatfelvételre van szükség.



2.9. ábra

*A lineáris és az adaptív technikával összeállított, azonos nehézségű modulokat tartalmazó tesztek információs függvényei (Forrás: Magyar és Molnár, 2015)*

## 2.5. A technológiaalapú tesztelésre való átállás előnyei, hátrányai és kihívásai

A technológia fejlődése és terjedése, általánossá válása, hozzáférhetősége rendkívüli lehetőségeket kínált és kínál a pedagógiai mérés-értékelés gyakorlatának fejlesztésére. Lényegében a számítógép oktatási célú alkalmazásával egy időben megjelent a számítógépes tesztelés, miután alkalmas a nagy hatékonysággal történő adatgyűjtésre. A feleletválasztós feladatokat minden nehézség nélkül át lehetett ültetni számítógépre, és ahogy a számítógépek fejlődtek, úgy alakultak ki az egyre fejlettebb számítógépes technikák. Ennek ellenére rendszeres, gyakorlatban történő alkalmazásuk relatív későn kezdődött, valójában ezek azok az évek, amikor a fejlődés világszerte felgyorsult és mára már nagy biztonsággal prognosztizálható, hogy belátható időn belül minden fontosabb mérés átkerül technológiai alapokra. A késői elterjedésnek előnye, hogy ma már gazdaságosan kihasználhatóak a technológiaalapú tesztelés előnyei (Csapó, 2014).

Alkalmazásának segítségével pontosabb, változatosabb, lényegesen komplexebb képességeket vizsgáló tesztelési eljárásokat, életszerűbb, alkalmazásorientáltabb, természetesebb környezeteket, feladatokat alakíthatunk ki a korábbi papíralapú mérésekhez képest (Bennett, 2002; Bridgeman, 2010; Farcot és Latour, 2009; Kikis, 2010; van Lent, 2010; Martin, 2010; Martin és Binkley, 2009; Moe, 2010; Ripley, 2010; Christakoudis, Androulakis és Zagouras, 2011; Csapó, Ainley, Bennett, Latour és Law, 2012; Beller, 2013; Breiter, Groß és Stauke, 2013). Ebben a fejezetben részletesen kifejtjük, hogy a technológiaalapú mérésre való átállás milyen előnyökkel, illetve hátrányokkal bírhat. Hatékonysága, a hatékonyság növekedése bizonyos feltételek mellett a mérés-értékelés minden egyes szintjén kimutatható:

- a) a tesztelés gazdaságossága (Farcot és Latour, 2008; Bennett, 2003; Choi és Tinkler, 2002; Peak, 2005; Rose, Hess, Hörhold, Brähler és Klapp, 1999; Wise és Plake, 1990);
- b) a tesztszerkesztés változatossága (Csapó és mtsai, 2012), a kiközzvetítés és adatáramlás gyorsasága (Csapó, Lőrincz és Molnár, 2012),

- c) az azonnali, objektív, standardizált visszacsatolás biztosításának lehetősége (*Becker, 2004; hosszabb, írott szöveges válaszok értékelése kapcsán: Dikli, 2006; Mitchell, Russel, Broomhead és Aldridge, 2002; Valenti, Neri és Cucchiarelli, 2003*),
- d) változik a diákok tesztelés iránt mutatott motivációja (*Meijer, 2010; Sim és Horton, 2005*),
- e) innovatív feladatszerkesztési lehetőségek, multimédiás, dinamikus, interaktív itemek, második és harmadik generációs tesztek alkalmazása (*Strain-Seymour, Way és Dolan, 2009; Pachler, Daly, Mor és Mellar, 2010*), amelyek korábban papíralapon nem voltak kivitelezhetőek (*Molnár, 2010a; Molnár, Greiff, Wüstenberg és Fischer, 2016*).
- f) Elérhetővé válik az adaptív tesztalgoritmus; amelynek segítségével pontosabbá válik a tudás- és képességszintbecslés (*Magyar és Molnár, 2013; Csapó, Molnár és R. Tóth, 2008; Frey, 2007; Jodoin, Zenisky és Hambleton, 2006; Magyar, 2012, 2013, 2014a, 2014b; Molnár, 2015a; 1. 2.4.3. részt*);
- g) bővül a tesztelésbe bevonhatók köre (pl. meghallgathatóak a feladatok; *Csapó, Molnár és R. Tóth, 2008; Molnár, 2015b*) és
- h) lehetővé válik a kontextuális adatok hatékony (pl. például mérhetjük az egyes itemek megoldásához szükséges időt, feltérképezhetjük hányszor javított a diák a megoldásán, hova és hányszor kattintott a tesztelés folyamán) rögzítése és elemzése is (*Csapó, Lőrincz és Molnár, 2012; Bodmann és Robinson, 2004*). Ennek következtében a papíralapú tesztelésnél megszokott egyedüli indikátor, a teszteredmény helyett gazdag és jól strukturált, a diák tesztelés alatt mutatott viselkedésének pontosabb követését lehetővé tevő adatbázis áll rendelkezésünkre (*Molnár és Lőrincz, 2012*).
- i) Javulhatnak a tesztek jóságmutatói (*Jurecka és Hartig, 2007; Ridgway és McCusker, 2003; Csapó, Molnár és Nagy, 2014, 2015*).

### 2.5.1. A tesztelés gazdaságossága, költségei

A számítógépes adatfelvétel egyik leggyakrabban említett előnye a papír-ceruza alapú teszteléssel szemben a hosszú távú költséghatékonyság (pl. *Bennett, 2003*). Annak ellenére, hogy a számítógépes tesztrendszer, valamint az adatfelvételhez szükséges infrastruktúra kiépítése költséges, a tesztrendszer üzembeállítása után annak rendszeres használatával a számítógépes mérés-értékelés számos megtakarítási lehetőséget kínál a papíralapú teszteléssel szemben.

A papíralapú tesztelés során minden egyes tesztelési alkalomkor elvégzendő tevékenységek a következők: feladatírás, lektorálás, feladatszerkesztés, tesztösszeállítás, nyomdai előkészítés, nyomtatás, csomagolás, szállítás, kiosztás, adatfelvétel, tesztek begyűjtése, szállítás, javítás, kódolás, adatok rögzítése, adattisztítás, teszteredmények kiszámolása, eredmények visszajelzése és tesztek tárolása. Minden egyes tevékenységhez hozzárendelhető annak költségvonzata.

A technológiaalapú, így a számítógépes adatfelvétel során sincs szükség a tesztek nyomtatására, fénymásolására, csomagolására, szállítására, válaszlapok készítésére, esetleg javításra, kódolásra, adatrögzítésre, automatizálható az adattisztítás, a teszteredmények kiszámolása, az eredmények visszajelzése és a tesztek tárolása, ezáltal az adatfelvétel költsége jelentősen csökken (l. *Choi és Tinkler, 2002; Paek, 2005; Rose és mta, 1999; Csapó, Molnár és R. Tóth, 2008; Wise és Plake, 1990; Bennett, 2003; Valenti, Neri és Cucchiarelli, 2003; Christakoudis, Androulakis és Zagouras, 2011; Csapó, Lőrincz és Molnár, 2012; Wang, 2004*).

A tesztelés módjának költségeit modellező elemzések szerint a dokumentációs költségek akár kétharmada spórolható meg (*Rose és mtsai, 1999*) számítógép-alapú tesztek alkalmazásával. A tesztelés után az esetleges kérdéses helyzetek kezelése esetén az elektronikusan rögzített adatok visszakeresése gazdaságosabban oldható meg, miután könnyebben megvalósítható, mint a nagy mennyiségű papíralapú dokumentációban történő visszakeresés.

A számítógép-alapú tesztelés gazdaságosságát jelentős mértékben növeli, ha a rendszer használatára nem csak egyszer, hanem minél gyakrabban sor kerül, ha ugyanazon platform, esetleg ugyanazon itemek többször is alkalmazásra kerülnek. Az előzetes feladatbankösszeállítási költségeken kívül a későbbiekben elhanyagolhatóvá válnak a rendszer használatának kiadásai.

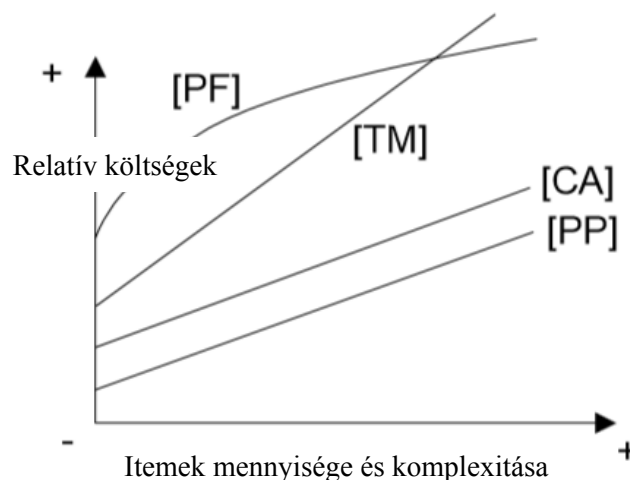
Miután a számítógépes adatfelvétel során a diákok válaszait a tesztelő rendszer automatikusan menti (*Bodmann és Robinson, 2004*; külső adathordozóra, számítógépre, vagy közvetlenül a központi szerverre) nincs szükség kódolók és adatrögzítők alkalmazására. Sőt összességében a feladatok javítási költségei is jelentős mértékben csökkennek, akár fel is számolhatóak a rendszeren belül, mert számos feladattípus javítása automatikusan megvalósítható, extra humán erőforrás bevonása nélkül.

Ahogy egyre nagyobb szerepet kap a tesztekben az automatikusan javítható feladatok aránya, úgy növekszik a teszt megbízhatósága, objektivitása, miután a számítógép automatikusan azonos kódolási, javítási módszert alkalmaz diák és iskolafüggetlenül. A tesztek eredményében más tényező (pl.: a javító szigorúsága) nem játszik szerepet. Ez szintén besorolható a gazdaságossági kérdések közé, miután megbízhatóbb és ezáltal gazdaságosabb tesztelési mód kialakítására van lehetőség. A kiértékelési megbízhatóság mértéke akkor maximális, ha az adatok közvetítő eszköz nélkül kerülnek rögtön az adatbázisba, a szerverre.

A számítógép-alapú tesztelés és a papíralapú tesztelés költségeinek, valamint a rendszerben lévő feladatokért kifizetett átlagos összeg nagysága jelentős mértékben függ a papíralapú tesztelés és a számítógép-alapú tesztelés technológiájától, az alkalmazott feladatok, itemek komplexitásától és mennyiségétől. A továbbiakban négy esetet különböztetünk meg (*Farcot és Latour, 2009* alapján):

- 1) papíralapú tesztelés (PP – *paper-and-pencil*);
- 2) számítógéppel segített tesztelés (CA – *computer-aided*): alapvetően papíralapú a tesztelés, de az itemek fejlesztése egy központi számítógépes rendszeren keresztül történik;
- 3) számítógép-alapú tesztelés a tesztelés számára kidolgozott rendszerrel (TM – *computer-based with taylor-made system*): mind az itemek fejlesztése, mind kiközvetítése, mind a válaszadás számítógépen történik. Az itemek végső formáját nem az itemírók szerkesztik, hanem programozók, speciális IT (*information technology*, információs technológia) szakemberek (pl.: OECD PISA 2009 ERA bizonyos részei vagy a PISA számítógép-alapú természettudományos műveltség teszt egyes itemeinek dinamikus értékelése);
- 4) számítógép-alapú tesztelés általános platform alkalmazásával (PF): az itemeket az előre kiképzett itemírók szerkesztik egy felhasználóbarát platformon keresztül (pl.: bizonyos részében az OECD PISA 2009 ERA felmérés vagy az SZTE OK Diagnosztikus mérések fejlesztése eDia kutatások).

*Farcot és Latour (2009)* költséghatékonyságot elemző modellje értelmében (2.10. ábra) egyértelműen többbe kerül a számítógép-alapú feladatok fejlesztési költsége, mint a papíralapúaké.



2.10. ábra

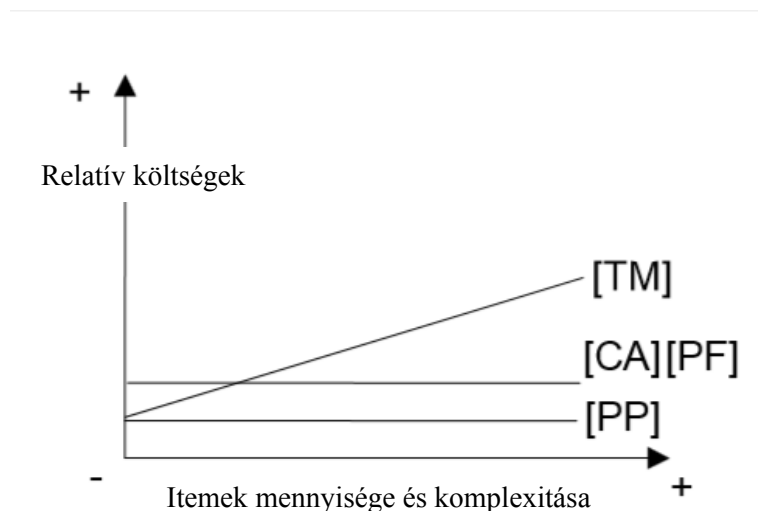
*Az itemfejlesztés költségei PP: papíralapú, CA: papíralapú, de számítógéppel segített, TM: számítógép-alapú informatikai szakemberek által programozott feladatok, PF: számítógép-alapú, feladatírók által szerkesztett feladatokkal (Forrás: Farcout és Latour, 2009. 113. o.)*

Az elektronikus feladatok előállítási költsége akkor kedvezőbb, ha a fejlesztések alapját egy általános platform képezi, amely segítségével feladatírók szerkesztik a feladatokat (természetesen többkörös lektorálással egybekötve, melynek költségeit mindegyik modell esetében azonos mértékben vettük figyelembe). E típusú megoldás kezdetben több képzés szükségességét vonja maga után, mint a programozók által készített feladatok, miután az előzetes képzésnek nem csak a feladatírásról, hanem az adott platform használatáról is szólni kell.

Az itemírás szervezeti költsége jelentős mértékben függ a használt technológiától. Papíralapú feladatok esetén a feladatírásért felelős csoport menedzsmentjének javadalmazása tartozik ide, amely egyrészt független a készített feladatok számától, másrészt függ a feladatok komplexitásától. Összességében egy fix paraméterű változóként szerepel a modellben. A CA scenárióban hasonló a helyzet, de az IT vonatkozás miatt egy előre meghatározott, rögzített összeg hozzáadódik a PP scenárióban számolt költségekhez. A TM scenárióban a külön programozott elemek kapcsán a külső erőforrások igénybe vételéhez szükséges szervezés többletköltséget jelent, míg a PF scenárió menedzselési költségei a CA scenárióval azonosak (2.11. ábra).

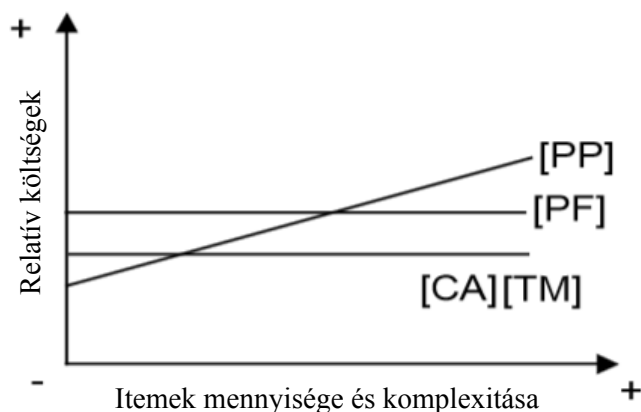
Mind a feladatírás, mind a tesztfejlesztés biztonsági kérdéseinek megoldása és a mérés tétjének megfelelő titkossági szint biztosítása is különböző költségekkel jár az említett scenáriók kapcsán. Az itemek mennyiségének és komplexitásának növekedésével, a bonyolultabb logisztika miatt a PP verzió már rövidtávon is jelentősen többbe kerül (kevés és egyszerű itemek alkalmazása esetén kedvező megoldás). A számítógéppel ötvözött scenáriók esetében a központi rendszer fizikális védelmének biztosítása független a rendszerben lévő itemek számától és komplexitásától, azaz azok nem befolyásolják e költségeket. A PF esetén a növekedett költség oka, hogy több ember sok helyről is csatlakozhat a rendszerhez, ezért az azonos szintű biztonság biztosítása többletköltséget jelent. Ha a csatlakozás helye jól definiált és centralizált, akkor a PF költségei azonosak a CA és TM modell esetén tapasztaltakkal (1.

2.12. ábra). Mindezen költségekhez hozzájárul a különböző helyiségek bérleti joga és az adott szcenárió működtetéséhez, fenntartásához szükséges infrastruktúra költségvonzata.



2.11. ábra

*Az itemírás szervezeti költségei (PP: papíralapú, CA: papíralapú, de számítógéppel segített, TM: számítógép-alapú IT szakemberek által programozott feladatok, PF: számítógép-alapú, feladatírók által szerkesztett feladatokkal) (Forrás: Farcout és Latour, 2009. 113. o.)*



2.12. ábra

*Az itemírás biztonsági költségei (PP: papíralapú, CA: papíralapú, de számítógéppel segített, TM: számítógép-alapú IT szakemberek által programozott feladatok, PF: számítógép-alapú, feladatírók által szerkesztett feladatokkal) (Forrás: Farcout és Latour, 2009. 114. o.)*

Összességében a PP tesztelés kiinduló költségei bizonyultak a legalacsonyabbnak, de ez a féle tesztelési mód csak abban az esetben marad hosszú távon is versenyképes, ha egyrészt nincs szükség sok feladat előállítására, másrészt az előállított feladatok komplexitása jellemzően alacsony. Ahogy nő az elkészítendő feladatok száma és nő a feladatok elvárt komplexitási szintje, fokozatosan úgy válik a PP tesztelési mód a legköltségesebbé a négy



scenário között. A TM és CA scénario kezdeti költsége magasabb, mint a PP-jé, azonban ha a feladatok komplexitása nem túlságosan magas, akkor már rövid távon is gazdaságosabb alkalmazásuk, mint a PP tesztelésé. Végül, bár a PF scénario alkalmazásának kezdeti költsége bizonyult a legmagasabbnak, mégis, már rövid távon is, ha sok, ráadásul komplexebb feladat kidolgozása a cél, a leg gazdaságosabb és leginkább fenntartható tesztelési módnak mutatkozott a négy scénario közül.

### 2.5.2. A kiközvetítés és adatáramlás gyorsasága, biztonsága

A számítógép-alapú tesztek kiközvetítése számos módon megvalósulhat (2.4. rész). Mindegyik kiközvetítési mód esetén különös figyelmet kell fordítani a biztonságra, meghatározott protokollok használatával akár növelhető is a papíralapú tesztelésnél tapasztalt biztonság. A számítógépes adatfelvétel során a diákok válaszait a tesztelő rendszer automatikusan menti (*Bodmann és Robinson, 2004*) a tesztelés típusától függően vagy egy pendrájvon, vagy a kliens számítógépére, vagy az iskolában/ vizsgaközpontban szerverként beállított számítógépre, vagy közvetlenül a központi szerverre, így nincs szükség például adatrögzítőkre, mint a papíralapú tesztelésnél. Ebből adódóan az adatok megbízhatósága is növekedik, mert egy fontos hibalehetőség kikerül a folyamatból. A megbízhatóság mértéke akkor maximális, ha az adatok közvetítő eszköz nélkül kerülnek rögtön az adatbázisba, a szerverre, azaz nem például egy pendrájvon keresztül, aminek esetleges meghibásodása adatvesztéshez is vezethet.

Az alkalmazott számítógép-alapú tesztelés és a használt mérés-értékelési rendszer típusától függően megvalósítható a feladatok azonnali javítása (kivételt képeznek a rajzolást, vagy hosszabb szöveg írását kérő feladatok), azaz az adatáramlás gyorsasága jelentős mértékben megnő. A számítógépes tesztelés nyújtotta automatikus kiértékeléssel szemben a papíralapú tesztelés során az item típusától függetlenül humán erőforrást igényel a válaszok javítása, rögzítése, ami jelentős mértékben lassítja az adatáramlás sebességét és rontja annak minőségét (lásd pl. *Butcher, 1987*).

Nemzetközi szinten vannak példák hosszabb, írott szöveges válaszok, esszék (pl.: *Dikli, 2006; Burstein, Chodorow és Leacock, 2004; Mitchell, Russel, Broomhead és Aldridge, 2002*) illetve szóbeli feladatok (*SpeechRaterTM – Bridgeman, Powers, Stone és Mollaun, 2012*) automatikus kiértékelésére is. Az értékelések alapját általában felszíni jegyek (helyesírás, nyelvtan, szöveg felépítése stb.) és nem a tartalom értékelése képezi. Ennek ellenére azok a kutatások, amelyek az e típusú számítógépes értékelés eredményét vetik össze humán bírálók értékelésével, magas korrelációs együtthatókról számolnak be. Ezek az alkalmazások azonban többnyire angol nyelvre készültek, várhatóan belátható időn belül magyar nyelvű alkalmazásuk nem történik meg.

A rendszer teljes működtetéséhez szükséges dokumentációs feladatokat automatizálva (*Rose és mtsai, 1999*) lehetővé válik az eredmények azonnali visszacsatolása is (*Choi és Tinkler, 2002; Noyes és Garland, 2008*), nemcsak a diákok, hanem igény szerint a tanárok, iskolaigazgatók, fenntartók számára (*Thompson és Weiss, 2009*). Az adatfelvételt és visszajelentést követően az elektronikusan rögzített adatok tárolása, illetve visszakeresése gazdaságosabb és könnyebben megvalósítható, nem igényel többletköltséget, mint a nagy mennyiségű papíralapú dokumentáció őrzése, szükség esetén visszakeresése.

### 2.5.3. Azonnali, objektív, standardizált visszacsatolás – előnyök és kihívások

Bármely mérés-értékelési folyamat, technika hatékonyságát jellemzi annak visszacsatolási ideje, a visszacsatolás gyorsasága, ami jelentős mértékben meghatározza a beavatkozás gyorsaságát, jelen esetben hatékonyságát is. Egyrészt minél rövidebb a visszajelzés ideje, annál hatékonyabb a visszajelentés, annál gyorsabb beavatkozást tesz lehetővé. Másrészt minél kevesebb az adatfelvételtől a visszacsatolásig végzett munkában az emberi munka, az emberi tényező, annál pontosabbá válik a tesztelés, annál kisebb a rendszerbe bekerülő hibák mennyisége.

A számítógépes rendszerre alapuló visszacsatolás előnye (1) a visszacsatolás típusától függetlenül kevesebb időt vesz igénybe elkészítése. Még folyószöveg típusú visszajelzés esetén is könnyebben alkalmazhatók a panelek, a korábban elkészített értékelések, melyek ismételt alkalmazása megvalósítható. (2) Az eredmények könnyen hozzáférhetővé tehetők mind a diákok, mind a pedagógusok, mind a döntéshozók, vagy a szülők számára (l. pl. Országos Kompetenciamérés vagy az eDia visszajelző rendszere; *Molnár, Papp, Makay és Ancsin*, 2015). (3) A számítógép-alapú teszteléssel összekötött számítógép-alapú visszacsatolás jelentősen lerövidíti a visszajelentési időt. Ez az alkalmazott itemek típusától függően (a javítás humán erőforrás-igényétől függően) változik. Ha kizárólag automatikusan javítható itemeket tartalmaz a teszt, akkor a teszt megoldása után a diákok azonnal láthatják teljesítményüket (l. pl. SZTE OK eDia rendszere; *Molnár és Csapó*, 2013; *Molnár* 2015a, 2015b). Természetesen az egész mintára vonatkozó elemzések, viszonyítási pontok rendszerben történő megjelenítése csak az egész minta tesztelése után valósulhat meg, esetlegesen a szakértői elemzések rendszerbe történő visszatöltése után. Erre azonban kész adatbázis esetén nem szükséges több hónap, sőt több hét sem. (4) Előzetesen beépített panelek segítségével részletes visszajelzés adható, a panelek és a visszajelzés is könnyen módosítható, átalakítható, szerkeszthető. (5) Az automatikus pontozás objektívebb, konzekvensebb értékelést valósít meg, biztosan mindenki ugyanazon értékrend, javítókulcs szerint értékelődik, függetlenül minden más egyéb tényezőtől (pl. az értékelő szigorúsága). Ez jelentősen növeli a teszt reliabilitását és validitását is (l. 2.5.4 részt; *Csapó, Molnár és Nagy*, 2014, 2015). (6) A számítógép-alapú visszacsatolás nehézsége a rendszer elkészítése, kiépítése, beüzemelése, ami jelentős mértékű többletmunkával jár.

### 2.5.4. A tesztek jószágmutatóinak javulása

A tesztek viselkedését, azaz a diákok teszten nyújtott teljesítményeinek általánosíthatóságát, a tesztek által tervezett és mért konstruktum azonosságát, valamint az adatfelvétel és a kiértékelés objektivitását a tesztek három jószágmutatója: a reliabilitás, validitás és objektivitás jellemzi. Utóbbi biztosítása a teszteredmények általánosíthatóságának és érvényességének is szükséges feltétele (*Rost*, 2004). A mérés objektivitása akkor biztosított, ha a tesztelt személy teszten nyújtott teljesítménye csak a saját tudás-, és képességszintjétől függ és lehetőség szerint minden más tényezőtől, mint például az adatfelvétel körülményeitől (adatfelvételi objektivitás), a válaszokat kiértékelő személy szigorúságától (kiértékelési objektivitás) független (*Csikós és B. Németh*, 2002).

Papíralapú tesztelés során az adatfelvételi objektivitást pontosan megfogalmazott adatfelvételre vonatkozó instrukciókkal érhetjük el, melyben meghatározzuk az adatfelvétel módját, körülményeit, a mérőbiztos által közölhető információk körét, és a teszt megoldása

során alkalmazható segédeszközöket (Csikos és B. Németh, 2002). Mindezek megvalósulása a kiértékelés során azonban nehezen ellenőrizhető, miután a tesztelés után kizárólag a tesztelt személyek által utoljára adott válaszok állnak rendelkezésünkre, azokról kell visszakövetkeztetni arra, mi történt, mi történhetett az adatfelvétel alatt, azaz valóban megvalósult-e az adatfelvételi objektivitás.

Technológiai környezetben is szükséges az adatfelvétel menetének, módjának pontos meghatározása, sőt kis téttel bíró tesztek esetén javasolt, nagy téttel bíró tesztek esetén nélkülözhetetlen a használt technológia teszteredményekre hatást gyakoroló paramétereinek megadása, mint például a javasolt monitorfelbontás. Időkorlátos feladatok alkalmazása során biztosított, hogy mindenkinek pontosan ugyanannyi idő álljon rendelkezésére az adott feladat, teszt megoldására. A papíralapú tesztekkel ellentétben az adatfelvételt megvalósító technológiai eszköz segítségével utólag is pontosan rekonstruálhatóvá tehető az adatfelvétel menete (l. részletesen 2.5.6. részt).

Az adatfelvétel objektivitásának növelése mellett például automatikus adatrögzítés és kiértékelés alkalmazásával növelhetjük a teszt kiértékelési objektivitását is. Az automatikus adatrögzítés segítségével kiküszöbölhetjük a figyelmetlenségből, elütésből származó rögzítési hibákat, továbbá a javító figyelmetlenségéből, szigorúságából, megítéléséből adódó eltéréseket, hibákat. A humán erőforrást igénylő feladatok javítása a számítógépes rendszerben történhet csoportosan is, azaz az azonos válaszok javítása egyszerre, azonos módon valósul meg, azonos értékelést kapva. Összességében, technológiaalapú tesztelés során megfigyelhető az adatfelvételi és az értékelési objektivitás növekedése (Csapó, Molnár és R. Tóth, 2008), aminek következtében nő a tesztek reliabilitása és validitása is (l. Csapó, Molnár és Nagy, 2014, 2015).

#### 2.5.5. Innovatív feladatszerkesztési lehetőségek

A technológiai környezet a korábbiaknál változatosabb feladatformák alkalmazását teszi lehetővé, több lehetőséget biztosítva innovatív itemtípusok használatára, mint a papíralapú tesztelés (Ridgway és McCusker, 2003; Jurecka és Hartig, 2007; Strain-Seymour, Way és Dolan, 2009; Csapó, Molnár és Nagy, 2014), ugyanakkor kihívást jelenthet néhány papíralapon bevált itemtípus számítógépen való megjelenítése (pl.: rajzolási feladatok).

Hangok, képek, animációk, videók, dinamikusan változó feladatok tehetik még színeesebbé, életszerűbbé a tesztek. A tesztelt személyek a feladatokra adott válaszaikat általában változatos formában adhatják meg, például kijelöléssel (szövegben szavak, mondatok kijelölése), kattintással (képek, alakzatok, szövegrészek), kattintással átszínezéssel, objektumok mozgatásával, átrendezésével (drag-and-drop típusú feladatok), összekötéssel (nyilak, egyenesek rajzolása két objektum közé), a feladat állítható feladatelemeinek bizonyos állapotba hozásával (pl. csúszka segítségével egy másik feladatelem bizonyos állapotának beállítása; Molnár, 2015c). A multimédiás szemléltető lehetőségek, amelyeket az innovatív feladatelemek kínálnak, ezen felül a munkamemória kapacitását is kevésbé terhelik, csökkentve a fáradtság okozta figyelemcsökkenés lehetőségét (Strain-Seymour, Way és Dolan, 2009).

Amellett, hogy az innovatív elemeket tartalmazó tesztek lényegesen realisztikusabb élményt okoznak megoldóiknak, az egyéb, vizuális vagy auditoros demonstrációs lehetőségeknek köszönhetően a tanulóknak kevesebb írott szöveget szükséges olvasniuk. Ezzel csökkentjük a lehetőségét annak, hogy egy gyengébb szövegértési képességgel rendelkező tanuló emiatt a hiányossága miatt egyéb képességterületen gyengébben teljesítsen, mint amire

valójában képes lenne. Technológiaalapú tesztek alkalmazásával még az olvasni nem tudó diákok is bevonhatóak a tesztelés folyamatába.

Az innovatív feladatszerkesztési lehetőségek köré sorolható az automatikus generálással előállítható (*Gierl és Haladyna, 2013*) és a feladat megoldása során dinamikusan változó feladattípusok köre is. Előbbi egyre nagyobb szerepet játszik azokon a területeken, ahol sok új, de a korábbi feladatok alternatívájának is tekinthető feladatra van szükség, miután egy értékelési rendszer összeállításának leginkább költséges része az itemfejlesztés. Az automatikusan generálható feladatokban egyes feladatelemek egy előre meghatározott tartományból (pl.: számok), halmazból (pl.: nevek, fogalmak, képek) választódnak ki előre meghatározott kritériumoknak megfelelően. Ennek következtében az eredeti feladat működéséhez közel álló feladatok hozhatóak létre, a változók számától és a halmazok méretétől függően akár nagy számban.

A feladatmegoldás során a feladattal történő interakció függvényében dinamikusan változó feladatok esetén a feladat kontextusa, a feladatban rendelkezésre álló információk annak függvényében változnak, hogy a tesztelt személy mit módosított, mit változtatott az adott problémán. Dinamikusan változó feladatok segítségével olyan képességterületek előzetes tudástól független mérése is megvalósítható, amely egyrészt kulcsfontosságúnak számít a 21. században, másrészt mérése-értékelése hagyományos technikák alkalmazásával nem megvalósítható. E feladattípus empirikus kutatás keretein belül történő alkalmazásával és a feladattípus által mérhetővé tehető konstruktum, az interaktív, dinamikusan változó problémakörnyezetben alkalmazott problémamegoldó képesség fejlődésével foglalkozik részletesen a 4.4 rész.

#### 2.5.6. Kontextuális adatok rögzítésének lehetősége

A technológiaalapú tesztelés nemcsak a válaszok rögzítésére biztosít lehetőséget, mint a papíralapú tesztelés, hanem az adatfelvétel során az adatfelvételt végző platformtól függően akár az összes interakció rögzítésére is, amelyek segítségével rekonstruálhatóak a tesztet megoldók egyéni tesztmegoldási folyamatai (*Tóth, Rölke és Goldhammer, 2012*). A válaszok és interakciók rögzítésére különböző adatbeviteli eszközök, technológiai megoldások állnak rendelkezésre. A hagyományosnak nevezhető, széles körben elérhető eszközök innovatív használatától (pl.: egér, billentyűzet, mikrofon, webkamera, érintőképernyő), a kognitív laboratóriumokban rendelkezésre álló speciális interfészeken keresztül (szemmozgáskövető speciális szemüveg) a kognitív idegtudományokban használt eszközökig végtelen a lehetőségek tárháza. A technológia segítségével rövid idő alatt nagy mennyiségű adat rögzíthetővé válik és az adatfelvétel végén nemcsak egy indikátor áll a kutatók rendelkezésére, hanem egy gazdag változórendszerrel jellemezhető adatbázis.

A rögzített logfájlok segítségével nyomon követhető, mi történt a teszt megoldása során. Mennyi időt töltött a tesztelt személy egy-egy feladat megoldásával (pl.: a több kognitív kapacitást igénybe vevő, tipikusan nehezebb feladatokon valóban hosszabb időt töltött-e el, vagy csak tippelt és továbbkattintott; *Bridgeman, 2010*); hányszor módosította válaszait, ha a teszt összeállítása lehetőséget teremtett rá; melyik feladatra tért vissza akár többször is; sőt, további eszközök (pl.: szem- és arcfigyelő technológia) alkalmazásával (*Molnár és Lőrincz, 2012*) még az is rekonstruálható, mikor hova fókuszált, milyen érzelmeket váltottak ki belőle a feladatok.

A kontextuális adatok rögzítése és elemzése jelentős mértékben hozzájárulhat (1) a vizsgált jelenség alaposabb megértéséhez, (2) a háttérben működő kognitív és affektív folyamatok feltérképezéséhez, (3) a teljesítményt befolyásoló tényezők magyarázatához, (4) a tesztelt személyek tesztmegoldási stratégiáinak azonosításához, (5) a kutatás reprodukálhatóságához, (6) valamint a feladatok, tesztek továbbfejlesztéséhez (Bodmann és Robinson, 2004; Csapó, Molnár és R. Tóth, 2008; Molnár, 2010a; Csapó, Lőrincz és Molnár, 2012; Molnár és Lőrincz, 2012; Jeni, Lőrincz, Nagy, Palotai, Sebők, Szabó és Takács, 2012; Yin, Chen, Sun, Worm és Reale, 2008).

A kontextuális adatok elemzése széles körben történő alkalmazásának gátja, hogy eddig csak korlátozott számban állnak rendelkezésre olyan elemzési eljárások, módszerek, melyek alkalmasak e típusú adatok feldolgozására (l. pl.: Mislevy, Behrens, Dicerbo és Levy, 2012; Goldhammer, Naumann, Stelter, Tóth, Rölke és Klieme, 2014). Új elemzési módszerek, eljárások, új tesztelmélet kidolgozására van szükség ahhoz, hogy a technológiaalapú tesztelés adta e lehetőség teljes mértékben kihasználásra kerülhessen.

#### 2.5.7. Adaptív tesztelés lehetősége

A személy tudás-, és képességszintjéhez illeszkedő adaptív tesztelés adta lehetőségekre, a megvalósíthatóság kihívásaira, az item és részteszt szinten megvalósuló adaptivitás adta előnyökre és hátrányokra a 2.4.3. részben tértünk ki részletesen. Az adaptív tesztelés megvalósíthatósága a technológiaalapú tesztelés egyik legjelentősebb előnye, mert segítségével kivitelezhető, hogy az adatfelvétel során a személyre szabott, a tesztelt személy számára egyénileg kiválasztott és közvetített feladatok többsége optimális kihívást biztosítson az egyén számára, amivel jelentős mértékben növelhető a motiváció, sőt flow (Csíkszentmihályi, 1997) élmény is elérhető a feladatok megoldása során, miközben jelentős mértékben megnő a tesztelés során kinyert információ mennyisége, pontossága.

#### 2.5.8. A tanulói motiváció változása

A számítógépes tesztelés lényegesen változatosabb, motiválóbb tesztkörnyezet kialakítását teszi lehetővé, amely sokkal inkább illeszkedik ahhoz az ingergazdag környezethez, amelyhez a mai diákok hozzászoktak (Choi és Tinkler, 2002; Sim és Horton, 2005; van Lent, 2010; Meijer, 2010; Csapó, Lőrincz és Molnár, 2012). A videó-, illetve online játékokon, számítógépen szocializálódott diákok hatékonyabban motiválhatóak olyan típusú tesztekkel, amelyek közelítenek a mindennapjaikban használt fórumokhoz, azaz színesek, dinamikusak, kevés szöveget, helyette sok képet, animációt tartalmaznak. A tanulók mind hazai (l. Molnár és Magyar, 2015), mind a nemzetközi vonatkozó kutatásokban (l. pl.: Martin, 2010) kifejezték preferenciájukat a számítógépes mérésekkel kapcsolatban. Érdekesebbnek, realisztikusabbnak tartják a számítógép-alapú teszteket, mint a papíralapúakat. Videón megtekintve könnyebb megérteni számukra például a természettudományos kísérleteket, mint pusztán egy leírás alapján, nem beszélve a feladaton belüli manipulálás lehetőségéről (Strain-Seymour, Way és Dolan, 2009).

Motivációs szempontból a legkomolyabb potenciált az innovatív itemtípusok és az optimális kihívást jelentő feladatok mutatják (Molnár, 2010a). A megkérdezett pedagógusok is ezekben az itemekben, az azonnali visszacsatolás lehetőségében illetve az adaptív tesztelés

megvalósíthatóságában (l. 2.4.3. részt) látják a számítógépes tesztelés tényleges jövőjét (*Molnár és Magyar, 2015*). Az optimális mértékű kihívás hiányában a tesztelt személy nem szembesül megfelelő mértékű kihívással, elveszíti érdeklődését, vagy túl nehéz feladat, túl nagy kihívást jelentő teszt esetén könnyen stresszes helyzetbe kerülhet, ami jelentős frusztrációt válthat ki belőle.

#### 2.5.9. Bővül a tesztelésbe bevonhatók köre

Az inkluzív nevelés általános elterjedésével együtt az inkluzív mérések igénye is megjelent a pedagógiai gyakorlatban. A technológiaalapú mérések számos olyan lehetőséget kínálnak, amelyek egyenlő esélyeket teremtenek a többségi és a speciális nevelési igényű (SNI) gyermekeknek. Az olvasási, illetve látási nehézséggel küzdő tanulóknak lehetőségük van az instrukciók akár többszöri meghallgatására. Az írásproblémákkal küzdő tanulók nem szenvednek hátrányt nehezen olvasható kézírásuk miatt, hiszen válaszaikat nem leírniuk, hanem begépelniük kell. A vak és gyengénlátó tanulók számára ma már speciális billentyűzetek érhetőek el a válaszadás segítésére, a gyengénlátók szabályozhatják a megjelenő szöveg betűtípusának méretét, emellett számos speciális eszköz (pl.: fejegér) alkalmazásával kezük mozgásában akadályozott tanulók is bevonhatók a tesztelés folyamatába. Mindezen túl kitágítható és megfelelő technológia alkalmazásával megvalósítható olvasni nem tudó és esetleg egeret, billentyűzetet még nem használó, korlátozott számítógépes tapasztalattal rendelkező kisgyermekek tesztelése is. Ez számos új lehetőséget terem az óvodáskori segítő-fejlesztő mérések terén, melyek eddig főképp egyéni adatfelvételen alapultak (*Csapó, Molnár és Nagy, 2014, 2015*).

Ebben a részben áttekintettük a technológiaalapú mérésekben rejlő többletlehetőségeket, előnyöket, hátrányokat, kihívásokat. Miután a gazdaság világában akármerre nézünk, mindenhol új technológia vesz minket körül – számítógépek, nyomtatók, szkennerek, pda-k, mobiltelefonok, plusz a mindezen eszközöket összekötő hálózatok (*Bennett, 2002*). Mindezen eszközök azonban nemcsak mindennapi életünkben, hanem a tanulás-tanítás folyamatát tekintve is számos új lehetőséget teremtenek. A megfelelően alkalmazott, fejlett technológia fokozhatja a tanítási-tanulási és mérés-értékelési folyamat hatékonyságát: a tanulók tudásának megkonstruálását, felmérését, esetleges hiányosságaik feltárását (*Csapó, Molnár és R. Tóth, 2008*).

A technológia mérés-értékelés folyamatába történő integrálása hatékonyabb tesztelést eredményez. Hasonlóan a mindennapi munkavégzéshez, a mérés-értékelés folyamatában is az információ elektronikus áramlása gyorsabb és olcsóbb (*Moss és Hendry, 2002*), mintha mindazt fizikailag, manipulatív tennénk (*Negroponte, 1995* idézi *Bennett, 2002*). A technológia olyan új lehetőségeket, új mérési módszereket tesz lehetővé, ami a hagyományos technikákkal nem megvalósítható (*Csapó, Ainley, Bennett, Latour és Law, 2012*). A hagyományosnak számító területek nagyobb pontossággal és hatékonysággal mérhetőek, továbbá olyan területek mérése-értékelése is megvalósíthatóvá válik, amire korábban nem volt lehetőség. E területek közé tartoznak azok a készség és képességterületek, ahol a technológia már az adott konstruktum meghatározása során alapvető szerepet tölt be (pl.: IKT műveltség, technológiailag gazdag környezetben történő problémamegoldás, digitális szövegek olvasása; l. *ACARA, 2012; Babad, Peer és Hobbs, 2012; Bawden, 2008; Becker, Hodge és Sepelyak, 2010; Bundy, 2010; Gamire*

és Pearson, 2006; Hobbs, 2011; Katz és Macklin, 2007; Koltay, 2009, 2010; Lévai, 2013; UNESCO, 2008; UNESCO és Microsoft, 2011; Tongori, 2012) vagy ahol a technológia a konstrukció mérésének eszköze (pl.: dinamika mérése, hálózati kapcsolatokon keresztül történő csoportmunka; I. *Computer Science Teachers Association Task Force*, 2011; Gerry, 2008; Hawkes, 2012; NAGB, 2013; *Partnership for 21st Century Skills*, 2008; Pearson *Educational Measurement*, 2003; Herzog, 2012; Herzog és Racskó, 2013; Zelman, Avdeeva, Shmis, Vasiliev és Froumin, 2011). Technológiaalapú tesztelés alkalmazásával kitágítható, szélesíthető a tesztelésbe bevont személyek köre is.

Ma már senki sem vitatja, hogy a technológiaalapú tesztelés rövid, vagy hosszabb távon kiszorítja a papíralapú tesztelést, ezzel forradalmasítva a mérés-értékelés célját és lehetőségeit is. Nem kétséges tehát, hogy a multimédiás alkalmazások új megvilágításba helyezik a tanulási és az azzal szorosan kapcsolatban lévő értékelés folyamatát, és lehetőséget nyújtanak az információ, illetve a tudás új, innovatív módon történő integrálására (Molnár, 2007). A modern IKT eszközök, webes alkalmazások új lehetőséget teremtenek az oktatás színvonalának és a mindennapi tanítási, értékelési gyakorlat minőségének növelésére. Ennek következtében mind hazai mind nemzetközi szinten jelentős erőfeszítések történtek a megvalósítás alapját képező platformok kidolgozása érdekében. A folyamatos fejlesztés alatt álló programok (pl.: TAO, eDia) egyesítik a papíralapú tesztelés kapcsán gyűjtött tapasztalatokat a technológia adta innovatív lehetőségekkel. Mindezzel párhuzamosan nagymértékben megnőtt a technológiaalapú tesztelés alkalmazásának aránya, fokozatosan átvéve a hagyományos papír, illetve szemtől szembe történő tesztelés szerepét, egészen a kis téttel bíró kismintás mérésektől a nagymintás nagy téttel bíró tesztelésig. Utóbbi esetén az egyöntetű átállás akadályozója az iskolai infrastruktúra sokszínűsége (hazai vonatkozásban l. részletesen Molnár és Pásztor-Kovács, 2015a; Hunya, Kőrösné dr. Mikis, Tarsayné Németh és Tibor, 2010, 2011), a pedagógusok egy részének attitűdje (hazai vonatkozásban l. részletesen Molnár és Magyar, 2015) és az, hogy világszinten kevés olyan mérés-értékelési platform érhető el, melyek nemcsak a feleletválasztós itemek, hanem a korábban említett innovatív feladat típusok kezelésére, esetlegesen azonnali kiértékelésére is képesek. Ennek ellenére várhatóan belátható időn belül minden fontosabb mérés átkerül technológiai alapokra. A rendszeres gyakorlati alkalmazás viszonylag késői megjelenésének előnye, hogy a technológiai fejlődésnek köszönhetően ma már a tesztelésben is valóban gazdaságosan ki lehet aknázni az előnyöket. Az innovatív feladatokból összeállított technológiaalapú tesztek validitása, ha azok a papíralapú tesztek felváltását valósítják meg, további kutatásokat igényel (l. 3. és 4. fejezet).

## 2.6. Technológiaalapú mérés-értékelés kialakulása, fejlődése, nemzetközi és hazai tendenciái

A technológiaalapú, azon belül is kiemelt szerepet játszó számítógép-alapú teszteléssel kapcsolatos kutatások már három évtizedes múlttal rendelkeznek. A három évtized alatt jelentős mértékben változott a technológia mérés-értékelésben betöltött szerepe és változtak a számítógép-alapú tesztelés lehetőségei. A kezdetekben – 90-es évek elején – még magán a technológián és az egyes eszközök adta mérés-értékelési lehetőségeken volt a hangsúly, kiemelt szerepet kapva a drága, kis számban, esetlegesen csak kognitív laborokban elérhető eszközöket. Ebből adódóan a technológiaalapú tesztelés kezdeti fázisában még nem a papíralapú tesztelés

egy lehetséges alternatívájaként, felváltójaként jelent meg a kutatásokban (*Baker és Mayer, 1999*), hanem egy új lehetőségeket kínáló, de alapvetően drága tesztelési módként.

Egy évtizeddel később, az ezredforduló körüli években, amikor rendszeressé váltak a nemzeti monitor-vizsgálatok és nemzetközi felmérések (pl.: *The National Assessment of Educational Progress – NAEP, Programme for International Student Assessment – PISA*), a technológiai eszközök, különös tekintettel a számítógép elterjedésével ismét a mérés-értékelés fókuszába került a technológiaalapú tesztelés. Ezen a ponton már kiemelt szerepet képviselt a széles körben rendelkezésre álló számítógépek segítségével történő adatfelvétel, mint a korábbi, papíralapú tesztelés felváltásának, a továbblépésnek egy lehetséges alternatívája (*Csapó, Ainley, Bennett, Latour és Law, 2012; OECD, 2010a; Bennett és mtsai, 1999; Bennett, 2010*). Ennek hatására a kapcsolódó kutatások alapvetően nem a technológia adta plusz lehetőségekre, hanem a médiahatás vizsgálatára koncentráltak. Ennek oka, hogy az USA-ban bizonyos nagymintás nemzeti méréseknek elszámoltathatóság szempontjából komoly tétje van, így a mérési eredmények érvényességén meglehetősen sok múlik. A papír és a számítógép-alapú tesztelés eredményeinek összehasonlító vizsgálatai (*Kingston, 2009; Wang, Jiao, Young, Brooks és Olson, 2008*) az elmúlt 15 év egyik legfontosabb mérés-értékelési témája volt (*Gallagher, Bridgeman és Cahalan, 2002; Bjerkestrand, 2009; Csapó, Ainley, Bennett, Latour és Law, 2012; Peak, 2005; Texas Education Agency, 2008*). Ma már jól dokumentált kutatási területnek számít.

Az összehasonlítás első lépcsőfokán főképp első generációs (papíralapon is megjeleníthető) tesztek alkalmazására került sor (*Pachler, Daly, Mor és Mellor, 2010*) és a papíralapú tesztek változtatás nélküli számítógépesítésével indultnak. A teszt itemeit, itemformáit változatlanul hagyva vizsgálták a közvetítő eszköz szerepét, összehasonlították a két különböző közvetítő eszközön megoldott tesztek eredményeit. Ebben az esetben a cél olyan számítógépes tesztek létrehozása volt, amelyek lehetőség szerint minél több elemében egyezik a papíralapú teszt formátumával, azaz a tesztelés menete továbbra is lineáris marad, a feladatok azonos sorrendben jelennek meg minden egyes tesztelt személy előtt (*Csapó, Molnár és R. Tóth, 2008*). Már ezen a szinten is számos kérdés vetődött fel, például számít-e, hogy pontosan ugyanazon feladatokat lássa a tesztelt személy a képernyőn, amiket papíralapon egy oldalon lát, vagy elegendő, ha egymás után jelennek meg ugyanazon feladatok, biztosítva a visszalépés lehetőségét. Vajon a teszt tulajdonságainak minimális megváltoztatása hatására még ugyanazt a tudást méri a papíralapú és a számítógép-alapú teszt, illetve meddig méri ugyanazt a tudást (*Molnár, 2008b*)? Összehasonlíthatóak-e a különböző médiumon felvett teszteredmények? Vajon milyen betűméret optimális a képernyőn, ami azonos hatást kelt az analóg papíralapú teszttel, mennyire befolyásolja a számítógépes teszt megoldójának eredményét számítógépes jártassága? A különböző számítógépes képességekkel rendelkező egyének teszten elért eredményét befolyásolja-e a médium megváltozása, a különböző médiumokon való feladatmegoldás különböző stratégiákat kíván-e (*Paek, 2005*), a médium változtatása hogyan hat a különböző korú, nemű, szocioökonómiai státusszal rendelkező egyénekre. Az eltérés különböző lehet egyes képességterületek, valamint azon belül az egyes item- és feladattípusok vonatkozásában.

A közvetítő eszköz szerepét vizsgáló kutatások jellemzően különböző feltételek mellett, különböző tulajdonságú mintákon és különböző mérőeszközökkel valósultak meg, ami jelentős részben nehezítette és nehezíti egy közös konklúzió kialakítását (*Bennett, 2003; Way, Davis és Fitzpatrick, 2006; Csapó, Molnár és R. Tóth, 2009; Wang, Jiao, Young, Brooks és Olson, 2008*;



Csapó, Molnár, Pap-Szigeti és R. Tóth, 2009; Kingston, 2009; Molnár, 2010a; Hülber és Molnár, 2013; Hülber, 2012). A kutatások közös sajátossága, hogy azok mintája jellemzően a tíz évnél idősebb, általában középiskolás (l. Bennett, 2003) diákok köréből kerül ki és kevés kutatás fókuszál kisiskolás diákok különböző tesztkörnyezetben (papíralapú, technológiaalapú, szemtől szembeni) mutatott viselkedésének jellemzésére, teljesítményének összehasonlítására (l. pl. Applegate, 1993; Hülber és Molnár, 2013). Ennek következtében minden egyes jelentős mérés-értékeléssel foglalkozó szervezet, intézet elindította saját kutatási feltételeihez igazodó, az áttérés biztosítását megfelelő mértékben támogató és segítő kutatásait (pl. Pearson Educational Measurement, 2003; Peak, 2005). E kutatássorozat hatására a korábbiaknál alaposabban dokumentált elemzések is publikálásra kerültek, ahol a diákok átlagos teljesítményének összehasonlításán túl, sor került egyéb változók, jellemzők szerinti médiahatáselemzésekre is [kor (Oregon Department of Education, 2007; Choi és Tinkler, 2002; Kingston, 2009); nem – l. 2.1. táblázat; szocioökonómiai státusz (SES; Horkay és mtsai, 2006; Poggio, Glasnapp, Yang és Poggio, 2005; Sandene, Horkay, Bennett, Allen, Braswell, Kaplan és Oranje, 2005); etnikum, számítógépes gyakorlottság (Pomplun, Ritchie és Cluster, 2006; MacCann, 2006; Blazer, 2010; Lissitz és Jiao, 2012; Poggio és mtsai, 2005); technológiai paraméterek (monitorméret, -felbontás; betűtípus, -méret; görgetés, visszalépés; Waters és Pommerich, 2007); a mért műveltségi terület (kontextus) (Kim és Huynh, 2006, 2007)].

Korábban önmagában az átlagok összehasonlítása nem fejezte ki, árnyalta megfelelően az esetek legnagyobb részében eltérő mintán felvett teszteredmények közötti különbségeket, valamint a különbségekre és azonosságokra fókuszáló elemzések feltételezték, hogy ugyanazon konstruktum mérésére került sor papír és számítógép-alapon. Ezt felismerve, illetve kihasználva a legújabb elemzési technikák adta lehetőségeket újabb kutatási kérdések megválaszolására is sor került kiemelt helyen kezelve a média konstruktumvaliditásra gyakorolt hatásának fontosságát (pl. Schroeders és Wilhelm, 2011; Martins, 2010; Shaw, Crisp és Johnson, 2012; Dolan, Burling, Harms, Strain-Seymour, Way és Rose, 2013; Blazek és Forbey, 2011; Lottridge, Nicewander, Schulz és Mitzel, 2010; Winter, 2010). A konstruktumvaliditás leginkább elfogadott elemzési módszere a strukturális egyenletekkel való modellezés (SEM; Choi és Tinkler, 2002; Sandene, Bennett, Braswell és Oranje, 2005; Pommerich, 2004; Poggio és mtsai, 2005; Johnson és Green, 2006; Bollen, 1989; Bollen és Curran, 2006).

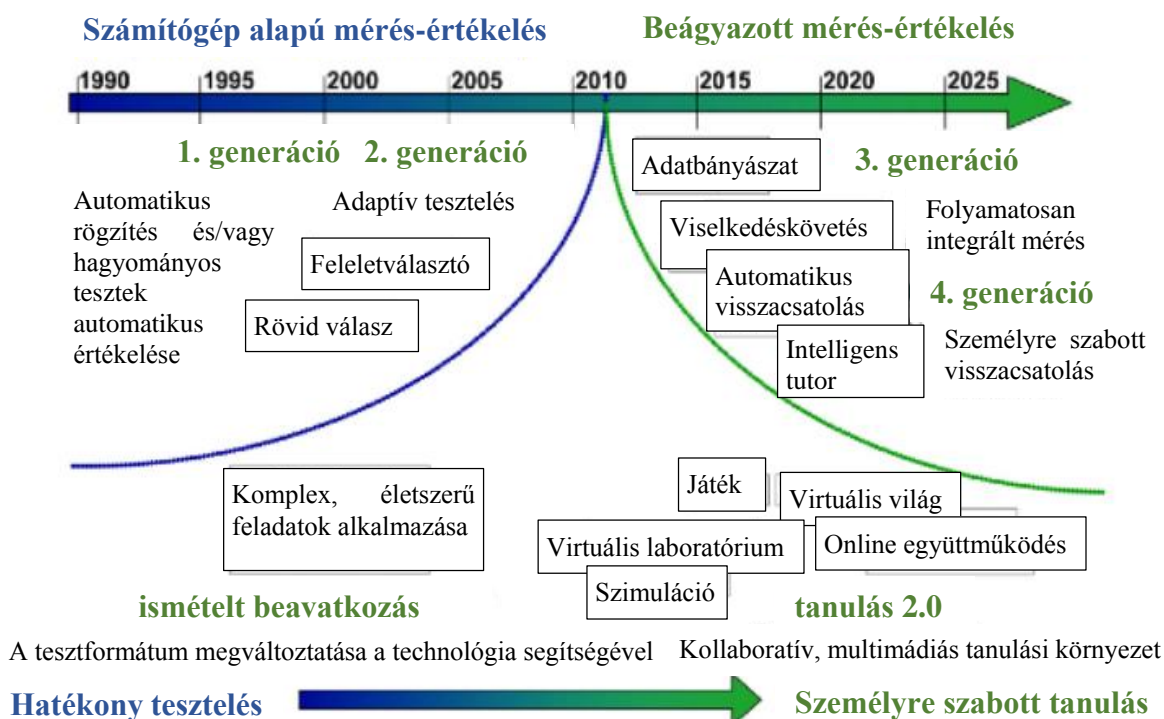
Összességében sem a nemzetközi, sem a hazai kutatások sem mutattak ki jelentős különbséget a nemek vonatkozásában. A tanulók életkorát és a közvetítő eszköz teljesítménybefolyásoló hatását görcső alá vevő kutatási eredmények az óvodáskorúaktól egyetemistákig vizsgálták a számítógépes tesztelés megvalósíthatóságát. Az eredmények biztatóak, bizonyos korlátozásokkal még a legfiatalabb korosztály tekintetében is megvalósítható a számítógép-alapú mérés. A médiahatás szocioökonómiai státusszal való kapcsolatának vizsgálata sem azonosított jelentős eltérést a kétféle médium vonatkozásában. Általánosságban megfogalmazható, hogy a közvetítő eszköz teljesítménybefolyásoló hatása nem szignifikáns, ha ugyanazon feladat papíron és monitoron történő megjelenítése közel azonos (pl. számítógépen nincs szükség görgetésre; Bridgeman, Lennon és Jackenthal, 2003), illetve a monitor felbontása, a feladat megjelenítése nem korlátozza annak olvashatóságát (Kingery és Furuta, 1997). A médiahatás annál inkább realizálódik, minél nagyobb az eltérés a kétféle médiumon történő feladatmegjelenítésben és a feladat feldolgozása és megoldása során szerepet játszó információfeldolgozó mechanizmusban.

2.1. táblázat. Kutatások, melyek a nem PP és CB teszteken nyújtott teljesítménybefolyásoló hatását vizsgálták

Publikáció	Minta (N)	Évfolyam	Képesség-terület	Kutatói elrendezés	A nem befolyásoló hatása
Taylor, Kirsch, Eignor és Jamieson (1999)	90000	nincs adat	olvasás-szövegértés	független minta	csekély mértékű
Clariana és Wallace (2002)	105	egyetemista	számítógép-használat	független minta	nincs
Wallace és Clariana (2005)	144	egyetemista	információs rendszerek	független minta	nincs
Gallagher és mtsai (2002)	773	nincs adat	matematika, olvasás-szövegértés	kevert elrendezés	nők esetén jelentősebb
Poggio és mtsai (2005)	2861	7.	matematika	független minta	nincs
Sim és Horton (2005)	20		természet-tudomány	független minta	nincs
Kies, Williams és Freund (2006)	384	egyetemista	orvos-tudomány	független minta	nincs
Bennett és mtsai (2008)	1970	8.	matematika	független minta	nincs
R. Tóth és Hódi (2011)	449	6.	olvasás-szövegértés	azonos minta	nincs
Moon (2013)	62953	2-8.	matematika	független minta	4. évfolyamon szignifikáns
R. Tóth (2015)	2217/ 449/ 342	5-7.	induktív gond./ olvasás-szövegértés/ komplex problémamegoldás	kevert elrendezés/ azonos minta/ független minta	nincs/ nincs/ nincs

Miután az utóbbi évtizedekben az oktatásban fokozatosan elterjedt a technológiai eszközök használata, a számítógép az iskolákban mindinkább alapfelszerelésnek tekinthető (Molnár és Pásztor-Kovács, 2015a; 1. 1.2. rész), fokozatosan háttérbe szorultak a közvetítő eszköz teljesítménybefolyásoló hatására fókuszáló kutatások (Way, Davis és Fitzpatrick, 2006) és megváltoztak a technológiaalapú mérések lehetőségei is (Mayrath, Clarke-Midura és Robinson, 2012b; Terzis és Economides, 2011). Előtérbe kerültek a korábban kihasználatlanul maradt lehetőségek (Greiff, Wüstenberg és Funke, 2012; Bennett, Persky, Weiss és Jenkins, 2007; Koong és Wu, 2011; Kuo és Wu, 2013), mint például különböző tudás és képességterületek új, innovatív eszközökkel, autentikus kontextusban történő vizsgálata.

Elképzelhetetlen lenne például a dinamikus problémamegoldó képesség vizsgálata statikus tesztkörnyezetben, a tesztváltozók manipulálásának lehetősége nélkül (Greiff, Wüstenberg és Funke, 2012; Molnár, Greiff, Wüstenberg és Fischer, 2016). Éppen így nem lehetséges a napjainkban eleminek számító digitális szövegek olvasási és értési képességének vizsgálata egy olyan tesztkörnyezetben, amely nem dinamikus, nem tartalmaz a weboldalakhoz hasonlóan megnyitható hiperlinkeket (Bennett, Goodman, Hessinger, Kahn, Ligget, Marshall és Zack, 1999; Bennett, Persky, Weiss és Jenkins, 2007; Koong és Wu, 2011; Kuo és Wu, 2013; Tóth és Hódi, 2013). A kollaboratív képességek mérése számítógép alapon szintén elképzelhetetlen lenne innovatív elemek, ez esetben a mérésbe épített chatfunkció biztosítása nélkül. A kommunikációs mintázatok rögzítése (Pásztor-Kovács, Magyar, Hülber, Pásztor és Tongori, 2013) már átvezet a legújabb kutatási tendenciákhoz, ahol már nem az egyén innovatív feladatkörnyezetben történő vizsgálatára helyezik a hangsúlyt, hanem a csoportfolyamatok mérésére is alkalmas kollaboratív mérési környezet kialakítására (pl.: *Assessment and Teaching of 21st Century Skills* (ATCS21), OECD PISA – kollaboratív problémamegoldás, *National Assessment & Testing*; Molnár és Pásztor-Kovács, 2015b; Pásztor-Kovács, 2016). Az értékelés azonnali visszacsatolásának lehetőségével élve a korábbi szummatív dominanciájú megközelítés mellett, előtt jelentős hangsúlyt kap a diagnosztika, az egyénre szabott, gyors, hatékony, tanulást segítő tesztelés, azaz a mérés tanulást segítő funkciójának kihasználása (2.13. ábra; Van der Kleij, Eggen, Timmers és Veldkamp, 2012; Kettler, 2011; Redecker és Johannessen, 2013; Molnár, 2015b).



2.13. ábra

A technológiaalapú tesztelés trendjei (Redecker és Johannessen, 2013 alapján)

A számítógépes tesztelés szélesebb körű alkalmazása a világ különböző régióiban az oktatás helyi sajátosságaitól függően más-más formában indult el (bővebben l. erről Molnár,

2010a; Csapó, Ainley, Bennett, Latour és Law, 2012). A leglassúbb fejlődés Európában, míg a leggyorsabb az USA-ban volt tapasztalható. Előbbi oka részben a tesztelési hagyományok hiánya, részben pedig a technológia lassúbb terjedése.

Európában az első országos és teljes körű számítógépes felmérésre Luxemburgban került sor, ami az országban tanuló diákok száma miatt bizonyos szempontból leegyszerűsítette a technológiai problémákat. A Luxemburgban kidolgozott TAO (a Test Assisté par Ordinateur rövidítése) nevű nyílt forráskódú tesztelő platformot (l. <http://www.taotesting.com/>) több ország is átvette. Az első németországi kutatások a DIPF-ben (a frankfurti központú Deutsche Institut für Internationale Pädagogische Forschung) is ezzel a programmal indultak el (<http://www.dipf.de/en/research/research-themes/technology-based-assessment>), és kezdetben a Szegedi Tudományegyetem Oktatáselméleti Kutatócsoportjának Diagnosztikus mérések fejlesztése projektje is a TAO platformot használta. Dániában szintén nagyobb léptékben folyik a technológiaalapú tesztelés, többi európai ország azonban csak a kísérletezés fázisánál tart.

Ázsia több országában jelentős szerepet kap, a tanterv részét képezi a technológia oktatási alkalmazása, ezért az értékelésben is megjelentek és rohamosan terjednek a különböző technológiai eszközök (UNESCO, 2014). Szingapúrban 1997-ben, Hong Kong-ban 2001-ben zajlott le vonatkozó átfogó tantervi reform, aminek hatására az infokommunikációs készségek fejlesztése meghatározó szerepet kapott az oktatásban. Japán e-tanulás stratégiája az 1990-es évek vége óta meghatározó és jelentős mértékben befolyásolja az oktatás fejlesztését is. A számítógépes oktatóprogramokba és a távoktatási rendszerekbe beépültek a tesztelő, visszacsatoló modulok, azonban a nagymintás nemzeti értékelési rendszerek még nem terjedtek el. A technológiaalapú tesztelés terén számos kutatási program indult el (Kampylis és mtsai, 2013).

Az Amerikai Egyesült Államokban nem csak az oktatásban, hanem az élet más területein (humán erőforrás menedzsment, katonaság) is jelentős hagyományai vannak a tesztek alkalmazásának. Feladatírásra, tesztfejlesztésre és azok alkalmazására üzleti vállalkozások sokasága épült. Amerikában terjedtek el először a tesztekre épülő elszámoltathatósági rendszerek (*test-based accountability systems*, l. Tóth, 2010, 2011), itt alkalmazták először a pedagógiai hozzáadott érték becslését (Balázsi és Zempléni, 2004), itt működik a világ legnagyobb pedagógiai értékelésre szakosodott kutató-fejlesztő központja (*Educational Testing Service*, ETS) és a 2001-es oktatási törvény is kiemelt szerepet szán a tanulók rendszeres felmérésének. A jelentős mérési hagyományok következtében természetesnek tekinthető, hogy Amerikában kezdődtek el legkorábban a számítógép-alapú tesztelésben rejlő lehetőségeket kihasználó kutatási-fejlesztési programok. Az ETS 1993 óta végzi a nemzeti monitorvizsgálatok szerepét betöltő NAEP (*National Assessment of Educational Progress*) felméréseket. Ezekhez a felmérésekhez kapcsolódóan zajlottak le az első átfogó technológiaalapú felmérések, amelyek középpontjában a matematika, az írás (Sandene, Horkay, Bennett, Allen, Braswell, Kaplan és Oranje, 2005) és a problémamegoldás állt (Bennett, Persky, Weiss és Jenkins, 2007). Jelenleg folyik a NAEP teljes átállása az online tesztelésre.

A nemzeti (szövetségi szintű) felméréseken túl az USA minden államában működik valamilyen szintű helyi mérési rendszer. Ezeknek az értékelési rendszereknek a lebonyolítói folyamatosan tanulmányozzák a technológia kínálta lehetőségeket, és különböző szintű kísérleteket folytatnak a technológiaalapú mérésekre való áttérés terén (Poggio és McJunkin, 2012). Az általános iskolai és a középiskolai szinten jelenleg a legismertebb és legnagyobb online felmérések, melyek főképp többszörös választáson alapuló feladatokra épülnek a MAP

(*Measures of Academic Progress*), a SOL (*Virginia Standards of Learning*) és az OAKS (*Oregon Assessment of Knowledge and Skills*). Az Obama adminisztráció hivatalba lépésekor (2009) meghirdette a *Race to the Top* (RTTT) oktatásfejlesztési programot. Ennek egyik komponense az új generációs tesztek kidolgozása. Jelenleg e program keretében zajlanak a világ legnagyobb költségvetésű online tesztelést fejlesztő munkái. A programban résztvevő államok két nagy konzorciumba tömörültek. A *Smarter Balanced Assessment Consortium* (SBAC) 30 állammal, a *Partnership for Assessment of Readiness for College and Careers* (PARCC) 25 állammal a közösen elfogadott sztenderdekre épülő online mérés-értékelési rendszert dolgoz ki. A SBAC nyílt forráskódú technológiát alkalmazva a legújabb online és adaptív vizsgarendszer kidolgozását célozta meg, míg a PARCC következő generációs, korábbiaknál jelentősen gyorsabb visszacsatolást biztosító, innovatív itemtípusok alkalmazását lehetővé tevő rendszer kidolgozását végzi. A cél, hogy mindkét mérési rendszer az új sztenderdek bevezetésével párhuzamosan a 2014/2015-ös tanévben már alkalmazásra kerüljön. A fejlesztések fontos hatása, hogy az amerikai tesztek nagy tétellel bíró (*high-stakes*) kontextusban fogják használni, így meg kellett oldaniuk az adatvédelem és adatbiztonság problémáit is.

#### 2.6.1. Technológiaalapú mérések a nemzetközi összehasonlító felmérésekben

Európában némileg később indultak meg a vonatkozó fejlesztések, de mára már jelentős szerepet kaptak a legprominensebb nemzetközi kutatásokban, az IEA és az OECD szervezet által koordinált vizsgálatokban. A számítógépes tesztelésre való átállás nehezíti a résztvevő országok sokfélesége és a technológia oktatásban játszott szerepének különbözősége. Az IEA által koordinált két fő program, a TIMSS (*Trends in International Mathematics and Science Study*) és a PIRLS (*Progress in International Reading Literacy Study*) egyaránt elindult a számítógép-alapú tesztelésre való átállás útján, az OECD fő programja, a PISA (*Programme for International Student Assessment*) már megtette az átállást, a szintén OECD szervezésű PIAAC (*Programme for the International Assessment of Adult Competencies*; OECD, 2013f) pedig már indulásakor eleve felkínálta a számítógép-alapú tesztfelvétel lehetőségét is.

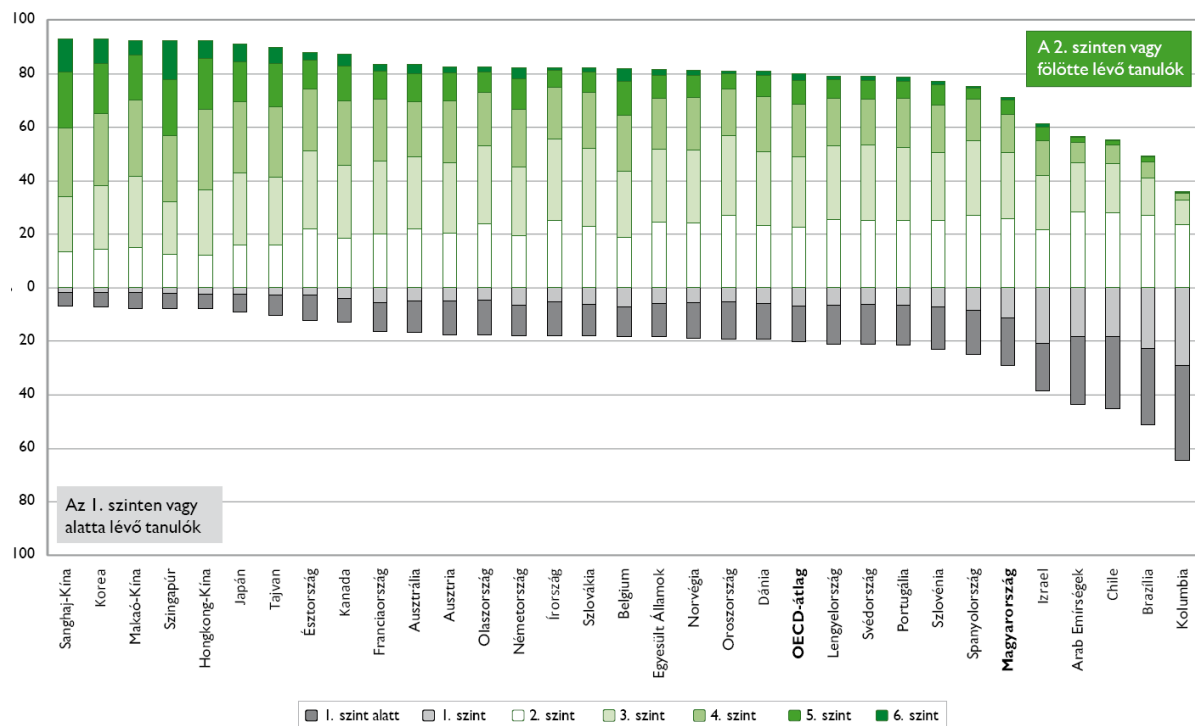
Az IEA-n belül a PIRLS 2016-os adatfelvétele során a papíralapú tesztelés lehetősége mellett a web alapú olvasás-szövegértés vizsgálata is megjelenik, miután 2013-ban az ICILS (*International Computer and Information Literacy*) kutatáson belül feltérképezte azt, hogy a tanulók miként használják a számítógépet például információszerzésre vagy kommunikációra (Fraillon, Ainley, Gebhart és Schulcz, 2013).

A sok éves előkészítő folyamatnak köszönhetően az átállás idejére az OECD PISA felmérései komoly számítógép-alapú hagyományokkal rendelkeztek. A 2000-ben induló felméréssorozatban a diákok IKT képességeinek kérdőíves feltérképezése (OECD, 2006) után 2006-ban jelent meg először technológiaalapú teszt, ekkor került sor a természettudományos műveltség számítógépes vizsgálatára. A résztvevő iskolákba hat, a tesztelés lebonyolításához szükséges programokkal ellátott laptopot szállítottak, amelyek közül egy a teszt adminisztrátori funkciókat látta el (Csapó, Ainley, Bennett, Latour és Law, 2012). A diákok válaszai mind az adott laptopon, mind az adminisztrátori laptopon rögzítésre kerültek. A számítógép-alapú természettudomány felmérésben (*Computer-Based Assessment of Science*, CBAS) még csak három ország vett részt, de az így kapott eredmények is fontos tapasztalatot jelentettek a következő fázisok megtervezéséhez (Moe, 2008; Sørensen és Andersen, 2009; OECD, 2009b).

A következő mérési ciklusban, 2009-ben szintén számítógép-alapú mérés, a digitális szövegértés szerepelt nemzetközi opcióként a programban (OECD, 2011). Ezek a tesztek nem a papíralapú szövegértés tesztek számítógépes megfelelőiként jelentek meg, hanem kifejezetten az vizsgálták, hogy a tanulók miképpen tudnak szövegeket értelmezni és felhasználni, információt gyűjteni számítógépes környezetben. A tesztelést nem az interneten keresztül hajtották végre, hanem pendrájvokat szállítottak ki az iskolákba, amelyek tartalmazták a tesztelés lebonyolításához szükséges programokat, illetve a diákok válaszait is a pendrájvon lévő adatbáziskezelő program kezelte, raktározta el. Ebben már 16 OECD ország, köztük Magyarország, és három további partnerország vett részt. A magyar tanulók eredménye a 19 részvevő között a 15 helyen állt. Ez a mérés már jelezte, hogy tanulóink digitális környezetben még gyengébben teljesítenek, mit a papíralapú teszteken.

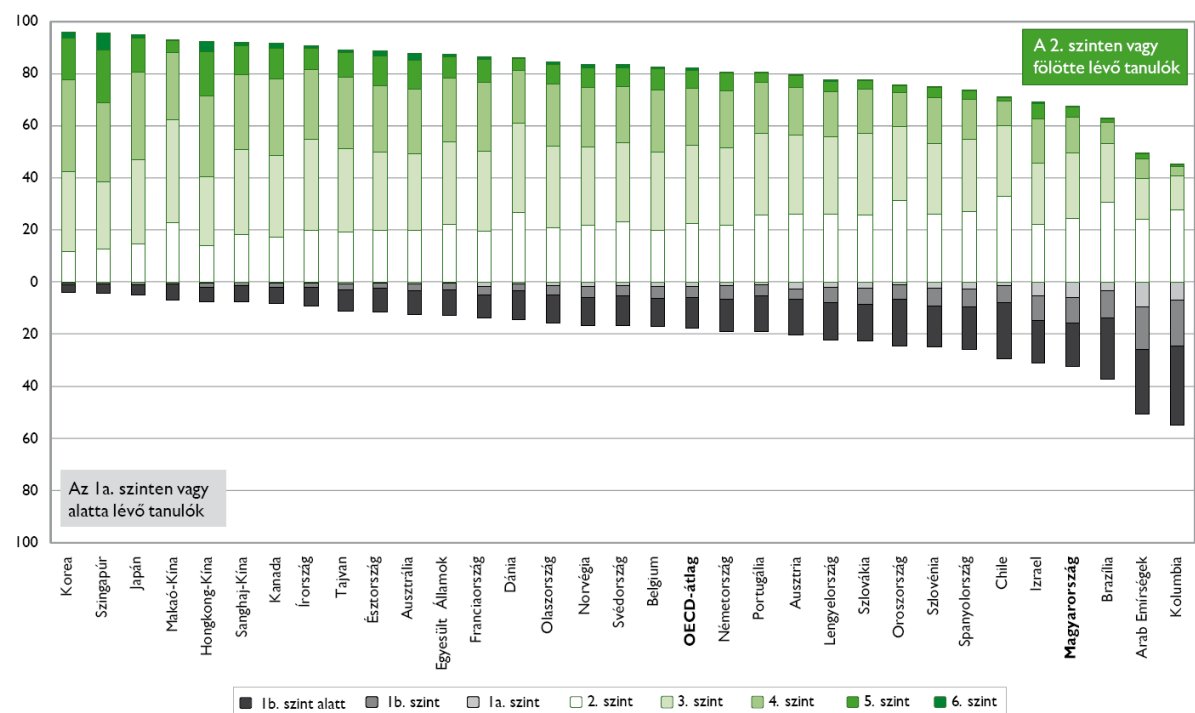
A 2012-es felmérés során a dinamikus problémamegoldás volt a negyedik, innovatív mérési terület, továbbá két fő területen (matematika és szövegértés) a papíralapú mérésekkel párhuzamosan számítógép-alapú mérésében is részt vehettek az arra vállalkozó országok. A párhuzamosan meghirdetett papíralapú és számítógépes felmérésnek az volt a célja, hogy pontosabban elemezzék a két médium különbségeit, a média-hatást, és eljárásokat dolgozzanak ki a kétféle mérés trendvonalainak összekapcsolására. A következő mérési ciklusban, 2015-ben alapértelmezésben már számítógépen zajlottak a mérések, és a papíralapú adatfelvételt csak azoknak az országoknak ajánlották fel, amelyek nem tudják teljesíteni a technológiai feltételeket. Magyarország 2012-ben is élt a technológiaalapú mérések lehetőségével, és mindhárom számítógépes adatfelvételben részt vett. A 2009-es mérésekhez hasonlóan ez a technológiai szempontból nagyobb kihívást jelentő felmérés is problémamentesen lezajlott. A magyar tanulók mindkét összehasonlítható esetben a papíralapú teszteken nyújtott teljesítmények alatt teljesítettek, és mind matematika, mind olvasás-szövegértés tekintetében a rangsor végére kerültek (l. 2.14. és 2.15. ábra; Balácsi, Ostorics, Szalay, Szepesi és Vadász, 2013).

A PISA 2012 felmérés negyedik, innovatív területe, a számítógépes környezetben megoldásra került kreatív problémamegoldás két fontos újdonságot is tartalmazott. Az egyik új elem a mérés tartalmi keretei által definiált problémamegoldó képesség (OECD, 2014). A teszt interaktív feladatainak megoldásához ugyanis nem volt szükség konkrét, az iskolában megszerezhető tárgyi tudásra. A problémák első felében a tanulók egy szimulált rendszerrel léptek interakcióba, ami során fel kellett ismerniük a rendszer működésének szabályait, majd azt ábrázolni egy sematikus modellben. A problémák második felében megismerhették a rendszer működése mögött rejlő valódi összefüggéseket, amelyeket ezután bizonyos előre meghatározott célok elérése érdekében alkalmazniuk kellett. Ez a mérés jól modellezte azt a gyorsan változó modern társadalmakban egyre gyakrabban előforduló helyzetet, hogy minden konkrét korábbi tudás nélkül kell újszerű problémákat megoldani, miközben csak az adott környezettel interakcióba lépve lehet szert tenni arra a tudásra, ami az adott probléma megoldásához szükséges. A problémamegoldás teszten a magyar tanulók a számítógépes főterületekhez hasonlóan szintén nagyon gyengén teljesítettek, 459 pontot értek el (az OECD országok átlaga 500 pont), a tanulók 35%-a nem érte el a kettes teljesítményszintet (a problémamegoldás egyáltalán értékelhető legalemibb szintjét).



2.14. ábra

*Az OECD PISA 2012 számítógépes matematika eredmények  
(forrás: Balázsi, Ostorics, Szalay, Szepesi és Vadász, 2013)*



2.15. ábra

*Az OECD PISA 2012 digitális szövegolvasás eredmények  
(forrás: Balázsi, Ostorics, Szalay, Szepesi és Vadász, 2013)*

Mindezen eredmények rávilágítottak arra, hogy Magyarországon már 2009-ben is megvoltak annak a technikai feltételei, hogy egy reprezentatív módon kiválasztott mintába bekerülő tanulók a nemzetközi normáknak megfelelő számítógépes tesztelésben vegyenek részt. Másrészt a gyenge eredmények utalnak arra is, hogy a tanulók iskolai számítógéphasználata messze nem olyan szintű, mint ami a fejlett országokban elvárható, és ami az OECD országok többségében meg is valósul. Erre utal a papíralapú és a számítógépes tesztek eredményei közötti nagy különbség. Harmadrészt, a tanulók teljesítménye a problémamegoldás teszten alatta volt annak a szintnek, amit a három fő mérési területen elért teljesítmények alapján várhattunk volna. Ez a tény pedig azt jelzi, hogy a magyar iskolák nagyon kevés figyelmet fordítanak a problémamegoldó gondolkodás fejlesztésére.

A modern társadalmakban a bonyolult problémák megoldása tipikusan csoportokban zajlik, ezért felmerült annak kérdése, miképpen készíti fel az iskola tanulóit az ilyen környezeten való helytállásra. A számítógépes értékelés legújabb kutatási tendenciái, beleértve az online értékelési platformok fejlesztését is, már nem az egyén innovatív feladatkörnyezetben történő vizsgálatára helyezik a hangsúlyt, hanem a csoportfolyamatok mérésére. Ez az ATC21S (*Assessment and Teaching of 21st Century Skills*) második fázisának egyik fő törekvése, és az OECD PISA 2015-ös mérésében is a kollaboratív problémamegoldás volt az innovatív mérési terület (OECD, 2013d). A csoportos problémamegoldás igen gyakran számítógépes hálózatokon keresztül zajlik, így az elvégzett műveletek és kommunikációs megnyilvánulások rögzítésével (logolásával) a problémamegoldás minden egyes eleme nyomon követhető és elemezhető. Ugyanakkor a kollaboratív problémamegoldás számos új kérdést is felvet, amelyek megválaszolása további kutatómunkát igényel. Az egyik alapvető mérésmethodikai újdonság az, hogy míg a mérések minden eddigi pszichológiai elmélete és teljes matematikai apparátusa az egyének mérésével foglalkozott, egyének megnyilvánulásaiból, teljesítményeiből következtetett pszichológiai sajátosságaikra, például képességeik szintjére, a kollaboratív problémamegoldás felmérése során csoportok teljesítményeiből kellene az egyének sajátosságaira következtetni. A másik nehézséget az okozza, hogy a valódi csoportmunka során meghatározó szerepe van az csoportdinamikának, a csoportok aktuális összetétele az egyének teljesítményeit is befolyásolja. E problémák áthidalására a PISA 2015-ös mérésben a csoportos munka egy szimulált rendszerben zajlott. Egy virtuális csoportnak csak egyetlen humán szereplője volt, a tesztet megoldó tanuló, a partnereket szimulált résztvevők, ügynökök (*agents*) helyettesítették.

Az OECD másik nagy vizsgálat-sorozata a felnőttek (16–64 év) képességeit méri fel. A PIAAC már indulása óta felkínálta a számítógép-alapú adatfelvételt. A PIAAC a felnőttek matematikai, szövegértési és problémamegoldó képességeit alapvetően technológiailag gazdag környezetben méri fel (OECD, 2013e, 2013f). A felmérés első fázisában egy személyes interjú-szituációban a vizsgálatvezető kérdéseket tett fel a felmért személyeknek, amelyeket hordozható számítógépen rögzített, ezután a tesztelt személy választhatott a papíralapú vagy a számítógép-alapú matematikai, illetve szövegértési tesztek között. A 2011-2012-es felmérésben részt vettek 74%-a a számítógép-alapú teszteket oldotta meg (OECD, 2013f). A problémamegoldó képességet vizsgáló teszt kizárólag technológialapon volt elérhető. Magyarország a PIAAC vizsgálatokban nem vett részt.

Összességében megállapítható, hogy a technológia használata két különböző célból jelenik meg a mérés-értékelés területén. A kutatások egyik felében azért alkalmaznak



különböző technológiai eszközöket, mert csak azokkal lehet felvenni az újonnan fejlesztett mérőeszközöket. Ezek a típusú felmérések drágák, ezért kis mintán, laboratóriumi körülmények között zajlanak általában. A kutatások másik fele a technológia nagymintás mérésekbe történő bevezetésére fókuszál, ahol a cél a papíralapú tesztelésről való áttérés. Ebből adódóan az alkalmazott technológiák terén inkább építenek a meglevő technológiai eszközökre (iskolai infrastruktúrára), vagy külön a mérés céljából szállítanak ki eszközöket (mint korábban láttuk ez azonban nagymintás mérések esetén jelentősen növeli a költségeket). A nemzetközi prominens mérés-értékelési projektekben nyújtott teljesítményeink egyöntetűen arra utalnak, hogy elodázhatatlan a számítógép-alapú mérés-értékelési kultúra széles körű elterjesztése és alkalmazása hazánkban, ha továbbra is jelen szeretnénk lenni az európai mezőnyben.

## 2.6.2. A számítógép-alapú tesztelés hazai fejlődése és tendenciái

A nemzetközi tendenciákat figyelemmel kísérve, követve és segítve a vonatkozó fejlesztéseket hazánkban is jelentős előrelépés történt a számítógép-alapú tesztelések területén, azonban e fejlesztések nem a nagy téttel bíró tesztek számítógépesítésére, hanem a diák és iskola szintjén is kis téttel bíró diagnosztikus értékelés irányába történtek. A Szegedi Tudományegyetem Oktatásméleti Kutatócsoportja egy uniós finanszírozású projekt keretein belül Diagnosztikus mérések fejlesztése címmel első lépésben adaptálta a PISA mérésekben is alkalmazott TAO nyílt forráskódú platformot (Csapó, Molnár és R. Tóth, 2009; Latour és Martin, 2007), majd elindította saját online értékelési rendszerének kidolgozását, az eDia fejlesztéseket (Molnár és Csapó, 2013; Molnár, 2015a, 2015b; Molnár, Papp, Makay és Ancsin, 2015; a projektről részletesebben l. edia.hu.). A program az online diagnosztika koncepciójának kidolgozásától a mérések tartalmi kereteinek fejlesztésén, egy elektronikus platform létrehozásán és feladatbakokkal való feltöltésén keresztül a gyakorlati kipróbálásig a teljes kutatási-fejlesztési és implementálási folyamatot átfogja (Steklács, Molnár és Csapó, 2015; Korom, Molnár és Csapó, 2015; Csíkos, Molnár és Csapó, 2015). Ennek következtében a diagnosztikus értékelés technológiai megvalósítása terén jelenleg Európában Magyarországon folyik a legjelentősebb fejlesztési program.

A formatív és a diagnosztikus értékelés számos ponton eltér a nagymintás szummatív tesztektől, így a diagnosztikus mérésekkel kapcsolatos eredményeket és tapasztalatokat nem minden tekintetben lehet közvetlenül átültetni a nagy téttel bíró nagymintás számítógép-alapú adatfelvételek lebonyolítására. Az egyik legfontosabb különbség az, hogy a diagnosztikus tesztek eredményeinek nincs messze ható, sorsdöntő következménye (*low-stakes test*), így nincs szükség a tesztmegoldás számítógépes környezetének standardizálására. A diagnosztikus tesztek hatékonyságát nem rontja lényegesen, ha minden tanuló a saját iskolájának épp rendelkezésre álló (iskolák között azonban esetleg nagyon különböző) számítógépén oldja meg a feladatokat, nem jelent kritikus problémát a sávszélesség, a szerver és a kliens oldal közötti kommunikáció sebessége, és az esetleg időlegesen fellépő kommunikációs problémák. Nagy téttel bíró teszt esetében azonban minden tanulónak megközelítőleg azonos feltételeket kell biztosítani a feladatok megoldásához. Ugyanezeknek a különbségeknek köszönhetően a diagnosztikus értékelés esetén nem jelent kritikus problémát a feladatok titkossága és az ehhez kapcsolódó adatbiztonság, továbbá a csalások megakadályozása sem. Mivel a formatív és diagnosztikus teszteknek csak a tanulás segítése a funkciója, a tanulónak nem érdeke, hogy problémáit eltitkolja, és az eredményeket úgy manipulálja, hogy azokban nem tükröződik a

valódi tudása. Vannak azonban a diagnosztikus és a szummatív teszt-rendszerek fejlesztése között olyan hasonlóságok, amelyek lehetővé teszik a diagnosztikus tesztek során szerzett tapasztalatok átvitelét a nagy téttel bíró tesztekre is. Ezek közé tartozik a tartalmi keretek kidolgozása, a számítógépes platform szerkezete, a kezelhető feladattípusok és az automatikus értékelés.

Hazánkban az első nagyobb volumenű online tesztelés 2008 májusában zajlott. Az adatfelvételben 5. évfolyamos diákok vettek részt ( $n=843$ ) 24 település, 34 iskolájából. A minta kiválasztása során nem volt cél a reprezentativitás. Ennek ellenére egy azonos korú reprezentatív és a jelen kutatás mintájának szülők iskolai végzettsége szerinti eloszlásában ( $\chi^2=7,13$ ,  $p>0,05$ ) nincs szignifikáns különbség.

A papíralapú teszt számítógépre adaptálása során a kutatók igyekeztek a teszt minél több tulajdonságát megőrizni (l. *Csapó, Molnár és R. Tóth*, 2009; *Csapó, Molnár, Pap-Szigeti és R. Tóth*, 2009). A kognitív teszt (induktív gondolkodás teszt) mellett mind a tesztelés előtt, mind utána egy-egy háttérkérdőívet (IKT kérdőív és szocio-ökonómia háttéradatakra kérdező kérdőív) is kitöltöttek a diákok, illetve a tesztelés során közreműködő tanárok is.

Minden egyes tanuló először papíron, majd néhány hét különbséggel számítógépen is megoldotta az induktív gondolkodás fejlettségét mérő tesztet. A teszt megoldására mindkét formátumban 35 perc állt a tanulók rendelkezésére. A számítógépes adatfelvétel interneten zajlott a TAO (*Testing Assisté par Ordinatur* – számítógépes tesztelés) platformon keresztül (*Plichart, Jadoul, Vandenabeele és Latour*, 2004; *Latour és Farcot*, 2008; *Latour és Martin*, 2007). A kutatás eredményeit l. *Csapó, Molnár és R. Tóth* (2009), illetve *Csapó, Molnár, Pap-Szigeti és R. Tóth* (2009) tanulmányaiban. Ez a típusú kutatás, amikor minden egyes diák mindkét médián megoldja az alkalmazott tesztet nemzetközi viszonylatban is egyedi. Ebben az esetben diákszinten összehasonlítható az alkalmazott médium befolyásoló hatása.

A fokozatos eltávolodás jegyében 2009 tavaszán az SZTE Oktatásméleti Kutatócsoport online tesztelés bevezetéséért felelős munkacsoportja folytatta a papír és számítógép-alapú tesztelés összehasonlító vizsgálatait. Három területen (matematika, olvasás és problémamegoldás) és két évfolyamon (2. és 6. évfolyam) hajtott végre párhuzamosan papír- és számítógép-alapú adatfelvételt. A minta és a mérőeszközök kiválasztása, illetve az alkalmazott itemtípusok különböztek az egyes területek esetében, így az elemzések során lehetővé válik a kutatás tervezéséből, az alkalmazott itemek különbözőségéből adódó eltérések pontosabb leírása, azonosítása (*Molnár*, 2009a; *R. Tóth*, 2009; *Hódi és R. Tóth*, 2009).

2010-ben folytatódtak a fent említett kutatások és a fent említett TAO platform egyik modulja segítségével, a TAO CAPI kérdőívmodullal megtörtént egy tanároknak szóló nem lineáris kérdőív felvétele (*Tóth*, 2009). A „Diagnosztikus mérések fejlesztése” program első fázisának tapasztalataira (l. *R. Tóth, Molnár, Thibaud és Csapó*, 2011) építve indult el, a 21. század oktatási, fejlesztési, mérés-értékelési igényekre építő új generációs pedagógiai mérés-értékelési platform, az eDia kidolgozása 2011-ben.

Az eDia (elektronikus diagnosztikus mérés-értékelési rendszer) kidolgozásának célja egy 1–6. évfolyamos diákok és pedagógusok által is használható, a tanítási, fejlesztési folyamat tervezését és megvalósítását segítő rendszer kialakítása (részletesen l. *edia.hu*). A rendszer tanórai, iskolai használatához elegendő egy általános böngésző és internetkapcsolat. Az eDia alapú tesztek változatos feladatformák alkalmazását teszik lehetővé, aminek következtében a feladatok utasításait nemcsak elolvashatják, hanem meg is hallgathatják a gyerekek, így a tesztek a még olvasni nem tudó vagy olvasási nehézségekkel küzdő diákok körében is

használhatók. A hangok mellett képek, animációk, videók teszik még színesebbé, életszerűbbé a feladatokat. A tanulók feladatokra adott válaszaikat változatos formában adhatják meg, például kijelöléssel, kattintással, átszínezéssel, mozgatással, átrendezéssel. Az eDia így nemcsak egyszerű feleletválasztós feladatok, hanem multimédiás elemek és a technológia új lehetőségeit maximálisan kihasználó feladattípusok közvetítésére is alkalmas. A rendszer a tesztelés végén azonnali visszajelzést ad a tanulók tudás- és képességszintjéről. A platform képes a személy képességszintjéhez illeszkedő tesztelés megvalósítására is, melynek feltétele a rendszer paraméterezett, indexelt feladatokkal való feltöltése. Mindezzel lehetőség nyílik egyrészt a már korábban is vizsgált tudás- és képességterületek új, innovatív és a diákok számára motiválóbb környezetben történő vizsgálatára (pl. nyelvi, zenei képességek), másrészt eddig nem vizsgált képességek jellemzőinek feltárására (pl. dinamikus problémamegoldó képesség), harmadrészt az objektív, viszonyítási pontokkal ellátott visszacsatolás biztosítására. A modern technológiát és a webes alkalmazásokat használó platform új távlatokat nyit az osztálytermi tevékenységek átalakításához és hatékonyan segíti a pedagógusok mindennapi oktatómunkáját, értékelési feladatainak ellátását (Molnár, 2015a, 2015b).

A Diagnosztikus mérések fejlesztése c. projekt keretein belül több mint 800 általános iskolában és több mint 100 középiskolában alkalmazták a pedagógusok a projekt keretein belül kidolgozott teszteket diákjaik képesség, és tudásszintjének meghatározására, az eDia rendszerrel történt adatfelvételi pontok száma meghaladja az egymilliót. A rendszer terhelhetőségét és stabilitását mutatja, hogy az adatfelvételek legsűrűbb napjain az oldallekérések száma is meghaladta az egymilliót.

### 2.6.3. A technológiaalapú mérés-értékeléssel kapcsolatos, a disszertációban felhasznált kutatásaink módszertanának áttekintése

A disszertációban 14 kutatás eredményére támaszkodom, melyeket (öt kivételével) a 2.2. táblázatban kronológiai sorrendben összegeztem. A kutatásba bevont vizsgálatok módszertana, mintajellemzője a vonatkozó fejezetekben részletesen kifejtésre kerül. A számítógép-alapú képességvizsgálatokat megelőző papíralapú kutatásaink mintáinak főbb jellemezőit a 4.1. táblázat foglalja össze, ezért a 2002 és 2011 közötti időszakban kivitelezett öt empirikus adatfelvétel közel 24.000 diákjára vonatkozó adatokat nem integráltuk a 2.2. táblázatba.

A kutatások felépítésének, a minta kiválasztásának, az adatfelvételi pont megválasztása során minden esetben figyelem előtt tartottuk, hogy a kutatások eredményei egyrészt alkalmasak legyenek nemzetközi szinten is új eredmények létrehozására, másrészt az azonos konstruktum mérését célzó vizsgálataink eredményei összeskálázhatóak legyenek, azaz biztosított legyen a megfelelő mértékű horgonyzás a különböző évfolyamokon és adatfelvételi pontokon keresztül is. Az egyes kutatások originalitására a vonatkozó fejezetben térek ki részletesen.

A vizsgált konstruktumok között kiemelkedő szerepet tölt be a problémamegoldó képesség. E képesség mérésnek változásán keresztül szemléltetem a technológia mérés-értékelés terén gyakorolt hatását, azokat a lehetőségeket, amelyek megvalósítása számítógép-alapú mérés-értékelés nélkül nem lenne lehetséges. A problémamegoldó képesség vizsgálata kezdetben hagyományos adatfelvételi technikákkal (első generációs tesztek), majd 21. századi adatfelvételi technikák alkalmazásával (második és harmadik generációs tesztek) valósult meg.

2.2. táblázat. A disszertációban bemutatásra kerülő empirikus kutatások listája és főbb jellemzői

Kutatás megnevezése	Adat-felvétel időpontja	Vizsgált jelenség	Elemszám (elemzésbe bevon)	Vizsgált életkori sáv
Papír- és számítógép-alapú tesztelés összehasonlító vizsgálata kisiskolás diákok körében	2010	induktív gondolkodás - médiahatás	PP: 952 CB: 377	1. évfolyam
A dinamikus problémamegoldó képesség fejlődése 5-11. évfolyamon	2011	dinamikus problémamegoldó képesség	855	5-11. évfolyam
Az iskolakezdés kulcsfontosságú előkészítéseinek online mérése kisiskolás korban	2012	beszédhanghallás, relációszókincs, elemi számlálás, tapasztalati összefüggés-megértés, tapasztalati következtetés	FF: 364/5423/ 1895/424/ 1094  CB: 364/1740/ 435/402/ 416	1. évfolyam
A dinamikus problémamegoldó képesség fejlődése 3-9. évfolyamon	2013	dinamikus problémamegoldó képesség	1291	3-9. évfolyam
Egér és billentyűzethasználati képességek fejlettségi szintje kisiskolás diákok körében	2013	egér és billentyűzethasználat	1195	1-4. évfolyam
Egér és billentyűzethasználati képességek fejlettségi szintje első évfolyamos diákok körében	2014	egér és billentyűzethasználat	4952	1. évfolyam
A dinamikus problémamegoldó képesség fejlettségi szintje 7. évfolyamon (longitudinális kutatások keretein belül kivitelezett adatfelvétel)	2014	dinamikus problémamegoldó képesség	1191	7. évfolyam
A dinamikus problémamegoldó képesség fejlődése 3-11. évfolyamon	2014	dinamikus problémamegoldó képesség	4371	3-11. évfolyam
A dinamikus problémamegoldó képesség fejlettségi szintje első évfolyamos egyetemisták körében	2015	dinamikus problémamegoldó képesség	1259	egyetem, 1. évfolyam (BA, osztatlan, nappali)

Mára a problémamegoldó képesség olyan dimenziójának mérése (dinamikus problémamegoldó képesség) is megvalósítható, melyre 15 évvel ezelőtt, hagyományos technikák alkalmazásával még nem volt lehetőség. A technológiaalapú mérést feltételező logfájelemzések és modern elemzési eljárások segítségével olyan kutatási kérdésekre is megtalálhatjuk a választ, amelyekre szintén nem volt lehetőség egy évtizeddel ezelőtt.

A 2.2. táblázatban felsorolt kutatások adatfelvételei alapvetően online adatfelvételek voltak, a kérdőívek, illetve tesztek kiközvetítését 2008-ig a TAO, majd a 2008 utáni

kutatásokban az eDia platform segítségével történtek. A számítógép-alapú adatfelvételek az iskolák számítógéptermeiben, az iskolák infrastruktúráját használva valósult meg. A tesztek, kérdőívek megoldására, kitöltésére egy tanóra, azaz 45 perc állt a diákok rendelkezésére.

Az elemzések során alkalmazott eljárások között szerepelnek klasszikus és valószínűségi tesztelméleti elemzések, valamint strukturális egyenletekkel történő elemzések. Az elemzések során az SPSS21.0, SigmaPlot12.0, MPlus6.0 és ConQuest programcsomagokat használtuk.

### 3. A TECHNOLÓGIAALAPÚ TESZTELÉS ALKALMAZHATÓSÁGA KISISKOLÁS KORBAN

Miután a jelentős, nemzetközi szummatív mérés-értékelési programok idősebb diákok tudás- és képességszintmérését célozzák, a technológiaalapú tesztelésre történő átállást segítő kutatások mintája is rendszerint idősebb korosztályokra fókuszálnak (l. 2. fejezet). Kisiskolás tanulók körében lényegesen kevesebb kutatás valósult meg (*Carson, Gillon és Boustead, 2011; Csapó, Molnár és Nagy, 2014, 2015*). Ugyanakkor, amíg idősebb diákok esetében feltételezhető a megfelelő szintű számítógép-használati jártasság, addig kisiskolás korban ez a kritérium nem feltétlenül adott.

Az óvodai, illetve az iskolába lépés időszakában megvalósuló technológiaalapú mérések esetében ezért elengedhetetlen annak vizsgálata, hogy a diákok képesek-e a különböző adatbeviteli eszközök adekvát használatára, felkészültek-e az online tesztekkel történő empirikus kutatásokra. A nem megfelelő szintű eszközhasználat komoly validitási problémákat eredményezhet, valamint a tesztelés frusztrálóvá válhat a gyerekek számára, és ellenérzés alakulhat ki bennük a technológiaalapú mérésekkel szemben.

A különböző számítógépes beviteli eszközök, például az egér és a billentyűzet kezelésének és a használat kora gyermekkori fejlődésének vizsgálata nem új kutatási terület, lényegében a számítógépek széles körű alkalmazásának elterjedésével a kutatók látókörébe került (l. pl. *Wilton és McLean, 1984; Crook, 1992; Lane és Ziviani, 2002*). Ezek a kutatások számos tanulsággal szolgálnak a technológiaalapú mérés-értékelés fiatalabb korosztályokra való kiterjesztéséhez, az itemek tervezéséhez és a tesztek összeállításához. Az eredmények azt mutatták, hogy az egérrel való kattintás már nagycsoportban (4–6 éves korban) sem okoz nagy problémát, ugyanakkor a sebességet és a pontosságot nagymértékben befolyásolja a kattintási terület mérete (*Donker és Reitsma, 2007a, 2007c; Hourcade, Bederson, Druin és Guimbretière, 2004*). Az egérhasználat kutatásának további területe a 'drag-and-drop', azaz a vonszolás műveletének vizsgálata. A kattintással szemben ennek a műveletnek a végzése már nehézséget jelenthet az iskolába lépő gyerekek tesztelése során. A pontosságot és a sebességet meghatározó tényezők közül a kutatások a vonszolandó és a célterület méretét, valamint a vonszolás irányát emelik ki: kisebb területek esetében, valamint a vertikális irányba történő vonszolás során az óvodás és iskolába lépő gyerekek több hibát ejtenek és lassabban hajtják végre a műveletet (*Donker és Reitsma, 2007b; Inkpen, 2001; Joiner, Messer, Light és Littleton, 1998*). A gyerekek gyakorlással gyorsan elsajátítják az egérhasználatához kapcsolódó műveleteket (*Lane és Ziviani, 2002, 2010*). A billentyűzet és az egér használatának összehasonlításakor az egér lényegesen hatékonyabb beviteli eszköznek bizonyult óvodások körében (4–5 éveseknél): egy navigációs feladatban a gyerekeknek gondot okozott a billentyűk azonosítása, valamint a figyelmük megosztása a billentyűzet és a monitor között (*Grünzweil és Haller, 2009*).

Az eredmények arra hívják fel a figyelmet, hogy körültekintően kell eljárunk a fiatal generációknak szánt alkalmazások és eszközök tervezése során, ugyanis a kisiskolás tanulók lassabban és több hibával hajtják végre a műveleteket, mint idősebb társaik (*Hourcade és mtsai, 2004*). Ezen megállapítás érvényes az iskolába lépő diákok technológiaalapú mérését lehetővé tevő szoftverekre és eszközökre is.

Az említett kutatások számos tanulsággal szolgálnak, ugyanakkor több olyan korlát is jellemzi őket, amelyek szükségessé teszik további vizsgálatok kivitelezését. A kutatások

különböző kontextusokban (például pszichomotoros fejlődés, szoftver- és eszköztervezés) vizsgálták a gyerekek eszközhasználatát, a célok között nem szerepelt egy standardizált teszt létrehozása. Jellemzően saját fejlesztésű szoftvereket és eszközöket használtak, a vizsgálatok kivitelezése más és más eljárások szerint zajlott, a felhasznált feladatok leírása során hiányoztak a tesztek jóságát jellemző megbízhatósági mutatók (*Lane és Ziviani, 2010*). A fentiekből következően az eddigi eredmények általánosíthatósága korlátozott, a megismételhetőségük nehézségekbe ütközik. További korlátnak tekinthető, hogy a kutatások többsége az 1990-es és a 2000-es évek elejéről származik, az utóbbi években kevesebb ilyen jellegű kutatás valósult meg. Ugyanakkor az utóbbi évtizedekre jellemző felgyorsult technológiai fejlődés nagy hatást gyakorolt az új generációkra, a generációs különbségek gyakran már néhány év különbséggel is megjelenhetnek.

Az eszközök tekintetében a vizsgálandó korosztályban kézenfekvő megoldásnak tűnhet a különböző érintőképernyős beviteli eszközök (pl. tablet, *Couse és Chen, 2010*) használata is, azonban ezen eszközök elérhetősége a hazai intézményekben jelenleg erősen korlátozott. Asztali számítógépek tekintetében iskoláink megfelelő számítógépes eszközparkkal rendelkeznek (*Molnár és Pásztor-Kovács, 2015a*), így amennyiben a diákok online tesztek kitöltésére való alkalmasságát szeretnénk vizsgálni, elsősorban az egér- és billentyűzet-használati képességeik fejlettségi szintjének feltérképezésére kell helyoznunk a hangsúlyt.

### **3.1. Az egér- és billentyűzethasználati képességek online mérése kisiskolás diákok körében**

#### **3.1.1. Célok**

A kutatás fő célja annak feltérképezése, hogy (1) egy egységes, vagy az egér- és a billentyűzethasználatot tekintve kétdimenziós, vagy műveletenkénti bontásban háromdimenziós konstruktnak tekinthetők-e az egér- és a billentyűzet-használati képességek; (2) alkalmas-e az iskolába lépő tanulók egér- és billentyűzet-használati képessége arra, hogy egyéb tudás- és képességszintjüket online mérjük; (3) vannak-e olyan egér- és/vagy billentyűzethasználatot igénylő műveletek, amelyek problémát jelentek számukra, ezért tesztbeli alkalmazásuk kerülendő; (4) a feladatokon, az egyes műveleteken nyújtott teljesítményüket időkorlát alkalmazása esetén mennyiben befolyásolja egér- és billentyűzet-kezelési képességeik fejlettségi szintje; (5) hogyan változik és miként jellemezhető 1–4. évfolyamos diákok egér- és billentyűzethasználati képességeinek fejlettségi szintje; (6) milyen háttérváltozók befolyásolják az egér- és billentyűzethasználatra vonatkozó képességeik fejlettségi szintjét.

A kutatás a vizsgált korosztály és konstruktnak tekintetében nemzetközi szinten is hiánypótló, ugyanis azok főképp idősebb korosztály tesztelésére fókuszálnak, feltételezve a megfelelő technológiai műveltség meglétét, illetve a papíralapú tesztelésről a számítógép-alapú tesztelésre való átállás vonatkozásában más képességterület kapcsán vizsgálják a technológia, a megváltozott közvetítő eszköz teljesítménybefolyásoló szerepét.

### 3.1.2. Módszerek

Az első, a kutatás során pilot mérésnek tekintett adatfelvétel mintáját 1-4. évfolyamos diákok (n=1195; l. 3.1. táblázat), míg a második, úgynevezett nagymintás adatfelvételét, kizárólag első évfolyamos diákok alkották (n=6962, l. 3.1. táblázat). A pilot kutatás mintájában minden egyes évfolyamon közel fele-fele arányban szerepeltek fiúk és lányok. A nagymintás adatfelvételben a lányok aránya 55% volt. A nagymintás adatfelvétel előtt az SZTE OK (Szegedi Tudományegyetem Oktatásméleti Kutatócsoport) több mint 600 partneriskolájának mérési koordinátorát kértük meg arra, hogy a 2014 szeptemberében iskolába lépő első évfolyamos diákokkal osztályszinten vegye fel az érintett tesztet. Az adatfelvételt nem vállaló iskolák az iskolai fülhallgató hiánya miatt nem vettek részt a tesztelésben, ugyanis kértük az iskolákat, hogy gondoskodjanak minden diák számára fülhallgatóról.

A mérőeszköz feladatai kialakításának elsődleges nehézsége volt, hogy ne tudás vagy egyéb képesség mérése valósuljon meg, hanem a diákok eszközhasználati sikerességének feltárása. Ennek érdekében a feladatok szövegei meghallgathatóak voltak, kizárva ezzel a diákok olvasási-képesség-szintjében lévő különbségeket és annak teszteredményekre gyakorolt hatását (l. pl. 3.1–3.4. ábra hangszóró ikonját).

*3.1. táblázat. A két adatfelvétel mintájának főbb jellemzői*

Adatfelvétel éve	Évfolyam	N	Nem (lányok aránya, %)
2013	1	406	51
	2	271	48
	3	220	49
	4	298	50
2014	1	4952	55

Minden feladatot egyszer kaphattak meg a tanulók (csak Tovább gombot tartalmaztak a feladatok), kizártuk a visszalépés és ezáltal a feladaton történő gyakorlás lehetőségét. Ha megszakadt az iskola internetkapcsolata, akkor ismételt belépés után annál a feladatnál folytathatták a megoldást, ahol abbahagyták azt.

Mind a pilot vizsgálatra kidolgozott, mind a pilot mérés eredményei alapján a nagymintás adatfelvételhez átdolgozott 44 ítemes teszt feladatai alapvetően két csoportba sorolhatók: (1) egérkezelést, (2) billentyűzetkezelést tartalmazó feladatok (l. pl. 3.1–3.4. ábra). A feladatok fokozatosan nehezedtek a teszten belül. Az egyes műveletek elvégzésének gyorsaságát nemcsak az idő logolásával, hanem időkorlátos feladatok alkalmazásával is mértük. Az egérkezelési feladatokban a kurzor/mutató megfelelő helyre történő navigálása mellett a kattintási, kijelölési műveletek pontosságát és a vonszolás technikájának elsajátítási szintjét is vizsgáltuk. Az egérkezelési feladatokat három részre bonthatjuk: (a) képelemeken történő navigálás és kattintás (pontosság, gyorsaság – 3.1. ábra); (b) űrlapelemeken (gomb, jelölőnégyzet, választógomb) történő navigálás és kattintás (pontosság, gyorsaság – 3.2. ábra); (c) vonszolás (drag-and-drop típusú feladatok; pontosság, gyorsaság – 3.3. ábra).





### 3.1. ábra

*Egérkezelést mérő példafeladat (képelemeken történő navigálás és kattintás – pontosság)*



### 3.2. ábra

*Egérkezelést mérő példafeladat (választógombra történő kattintás – pontosság és gyorsaság)*

A billentyűzetkezelés vizsgálata első sorban a gépelés funkció kiemelésével történ. A leírandó szövegek tulajdonságait alapvetően három jellemző szerint csoportosíthatjuk: (1) a szövegek hossza, (2) a szövegek értelmessége (esetlegesen idegen szavak előfordulása), valamint (3) a leírandó szövegben lévő különleges karakterek száma. A pilot kutatásban

szereplő teszt feladatai között ezek minden egyes kombinációja szerepelt, plusz minden egyes kombináció időkorláttal és időkorlát nélkül is helyet kapott a teszt feladatai között. Természetesen a szövegdobozba történő íráshoz a diákoknak egerhasználatra is volt szüksége, miután első lépésként bele kellett kattintaniuk a válaszmezőbe. A pilot mérés eredményeire támaszkodva az átdolgozott gépelési képességek fejlettségi szintjét mérő feladatok kizárólag egy-két (nem különleges) karakter begépelését kérték. A feladatok kialakítása során törekedtünk arra is, hogy a begépelendő betű kisbetűs és nagybetűs alakja lehetőség szerint azonos legyen, azaz a feladat megoldása során ne legyen szükség a betűk ismeretére, elegendő legyen a monitoron megjelenő alakzat és a billentyűzeten szereplő betűk mint alakzatok összevetése. Miután a nagymintás adatfelvétel szeptemberben történt, közvetlenül az iskolába lépés után, amikor még nem várható el, hogy a diákok ismerjék a betűket, azok kisbetűs és nagybetűs változatát, ezért volt szükség a begépelendő betűk körültekintőbb kiválasztására. A szövegdobozba történő íráshoz a diákoknak egerhasználatra is volt szüksége, miután első lépésként bele kellett kattintaniuk a válaszmezőbe. Ennek módját a billentyűzet-használati képességeket mérő részteszt első feladatában magyaráztuk el a diákoknak. A későbbi feladatokban (pl. a 3.4. ábra mintafeladatában) erre már nem került sor.



3.3. ábra

*Egerkezelést mérő példafeladat (vonszolás pontossága)*

A kutatási kérdések megválaszolásához alkalmazott elemzések között szerepelnek klasszikus tesztelméleti, valószínűségi tesztelméleti és strukturális egyenleteken nyugvó eljárások is. A valószínűségi tesztelméleti elemzések alapját a Rasch modell képezte, az elemzéseket a ConQuest szoftverrel végeztük. Az 1–4. évfolyamos diákok eredményeinek közös képességskálára transzformálását a Rasch modellel végeztük, a diákok képességszintjének meghatározásához 'mle' (Maximum Likelihood Estimate) értékeket, míg az egy kohorszra vonatkozó átlagos képességszint meghatározásokhoz plauzibilis értékeket (pv) számoltunk.

A konstruktum dimenzionalitását vizsgáló strukturális egyenletekkel történő elemzéseket az Mplus program segítségével végeztük el. Mivel a teszt itemei – kivétel nélkül – dichotóm itemek voltak, a modellillesztés során WLSMV (*Weighted least squares mean and variance adjusted*) közelítési eljárásra és THETA parametrizáció használatára került sor (Muthén és Muthén, 2010). A modellilleszkedés-vizsgálatok során  $\chi^2$  illeszkedésvizsgálatot, valamint a CFI (*Comparative Fit Index*), a TLI (*Tucker-Lewis Index*) és az RMSEA (*root mean square error approximation*) illeszkedésmutatókat használtuk. A modell által megmagyarázott teljes varianciát jellemzi a CFI és a TLI index, 0,90 érték feletti szintjük kívánatos (Bentler, 1990). A rezidumok varianciájára vonatkozik az RMSEA értéke, melynek elvárt értéke kisebb, mint 0,08 (Browne és Cudeck, 1993; Fan és Sivo, 2005; Vandenberg és Lance, 2000). Egymásba ágyazott modellek összehasonlítására nem alkalmazhattuk a tradicionális  $\chi^2$  különbségtesztet, helyette az MPlus speciális DIFFTEST eljárását, egy speciális  $\chi^2$ -próbát használtunk (Muthén és Muthén, 2010).



3.4. ábra

*Billentyűzetkezelést mérő példafeladat (gépelés pontossága)*

### 3.1.3. Eredmények

A pilot mérés eredményét egyrészt alátámasztotta, másrészt finomabb felbontású elemzésekre adott lehetőséget a nagymintás kutatás kizárólagosan első évfolyamos diákokra fókuszáló adatfelvételének elemzése. Ezért az eredmények ismertetése során az első négy kutatási kérdés kapcsán a nagymintás, a tágabb életkori intervallumot és speciális háttérváltozókat igénylő 5-6 kutatási kérdés kapcsán – az érintett háttérváltozókra vonatkozó adatok nem álltak rendelkezésünkre a nagymintás adatfelvétel adatbázisában – pedig a pilot kutatás eredményeire támaszkodunk. Utóbbi részletes elemzése Molnár, Tongori és Pluhár (2015) tanulmányában olvasható.

A pilot adatfelvétel 41 feladatának feladatbank szintű reliabilitásmutatója (személyszeparációs reliabilitás=0,77) és az azokból összeállított tesztek belső konzisztenciája megfelelőnek bizonyult. A nagymintás adatfelvétel 44 itemből álló tesztjének reliabilitásmutatója (Cronbach- $\alpha$ ) 0,89 volt, azaz a teszt alkalmasnak bizonyult a diákok egér és billentyűzethasználati képességei fejlettségi szintjének meghatározására.

A vizsgált konstruktum dimenzionalitásának vizsgálatára irányuló elemzések alapján a kétdimenziós modell, ahol külön dimenzióba soroltuk az egér- és a billentyűzet-használati készségeket, szignifikánsan jobban illeszkedett az adatokra, mint az egydimenziós modell ( $\chi^2=1387,871$ ,  $df=1$ ,  $p<0,001$ ). Általánosságban a kétdimenziós mérési modell illeszkedése megfelelő volt (l. CFI, TLI, RMSEA értékeket 3.2. táblázat), ugyanakkor az illeszkedési mutatók szignifikánsan javultak, amikor az elvégzendő művelet szintjén további dimenzió bevonására került sor. A háromdimenziós mérési modell, ahol külön dimenzióba került az egérhasználati készségeken belül a kattintás és vonzolás művelete, még pontosabban illeszkedett az adatokhoz ( $\chi^2=201,313$ ,  $df=2$ ,  $p<0,001$ ), ezért műveleti szinten, dimenziókénti bontásban mutatjuk be az eredményeket.

A feladatok nehézségi szint tekintetében megfelelőek voltak az iskolába lépő diákok számára, bár a képességskála lefedése nem volt teljesen egyenletes. Az alapvető egér- és billentyűzethasználatot igénylő műveletekre építő feladatok jelentős része könnyű volt az első évfolyamos diákoknak. Ezt mutatja a 3.5. ábra többdimenziós személy-item térképe.

A diákok képességszint szerinti elhelyezkedése és a feladatszámok nehézség szerinti elhelyezkedése nem teljesen párhuzamos, a diákok többségének a tesztben előforduló feladatok jelentős része nem jelentett kihívást, azokat több mint 50% valószínűség mellett jól meg tudták oldani. A teszthez modellált ideális populáció átlagos képességszintje a 0 logitegység körül lenne, jelen esetben, jelen esetben, dimenziótól függetlenül, jelentős mértékben 1 logitegységgel meghaladja azt.

*3.2. táblázat. A mérési modell illeszkedésmutatóinak jósága, a billentyűzet és egérhasználati képességek dimenzionalitás-vizsgálata*

Modell	$\chi^2$	$df$	$p$	CFI	TLI	RMSEA (95% CI)
Egydimenziós	13064,76	527	0,001	0,806	0,793	0,058 (0,058–0,059)
Kétdimenziós	5963,85	526	0,001	0,916	0,910	0,039 (0,038–0,039)
Háromdimenziós	5569,46	524	0,001	0,922	0,916	0,037 (0,036–0,038)

Megjegyzés: Kétdimenziós: egér- és billentyűzet-használati készségek külön dimenzióban, Háromdimenziós: a kattintás, a gépelés és a vonzolás műveletei külön dimenzióban.

A nagymintás adatfelvételben részt vevő diákok negyedének teljes teszten nyújtott teljesítménye 80%-os vagy e fölötti volt, mindössze 15%-uk átlagteljesítménye bizonyult 50% alattinak. A teszt 10 feladatát (a személy-item térkép jobb oldalán legalul lévő feladatok), melyek között kizárólag egérhasználatot igénylő feladatok szerepeltek (nagy és kis, egy vagy több objektumra való kattintás, illetve vonzolás), még a legalacsonyabb képességszintű diákok

is több mint 50% valószínűség mellett oldották meg helyesen. A legjelentősebb képességszintbeli különbségek a gépelés típusú feladatok kapcsán realizálódtak.

	Kattintás	Vonszolás	Gépelés	+item
6				
			X	
			X	
5			X	
			X	
4			XX	
	X		XX	
	X	X	XXX	
	XX	X	XXX	36
3	XXX	X	XXXX	
	XXXXX	XX	XXXXX	
	XXXXXX	XXXX	XXXXXX	27
	XXXXXXXX	XXXXXXXX	XXXXXXXX	14 15 38
2	XXXXXXXXXX	XXXXXXXX	XXXXXXXX	12 37
	XXXXXXXXXXXXX	XXXXXXXXXX	XXXXXXXX	9 22 23
	XXXXXXXXXXXXX	XXXXXXXXXX	XXXXXXXX	
	XXXXXXXXXXXXX	XXXXXXXXXXXXX	XXXXXXXX	26 31
1	XXXXXXXXXXXXX	XXXXXXXXXXXXX	XXXXXXXX	45
	XXXXXXXXXXXXX	XXXXXXXXXXXXX	XXXXXXXX	34
	XXXXXXXXXXXXX	XXXXXXXXXXXXX	XXXXXXXX	13 32 44
	XXXXXXXXXX	XXXXXXXXXXXXX	XXXXXX	25 40
0	XXXXX	XXXXXXXXXX	XXXXXX	8 21 41 43
	XXXX	XXXXXXXXXX	XXX	2 11 24 35 39
	XXX	XXXXX	XXXX	7 19
-1	XXX	XXXX	XXX	16 18 28
	X	XX	XXX	4 30 42
	X	XX	XXX	3 33
	X	X	XX	10 29
-2	X	XX	XX	1 5 17
	X	X	XX	6 20
			X	
-3			X	
			X	
			X	
			X	
-4			X	
-5				
			X	
-6				

3.5. ábra

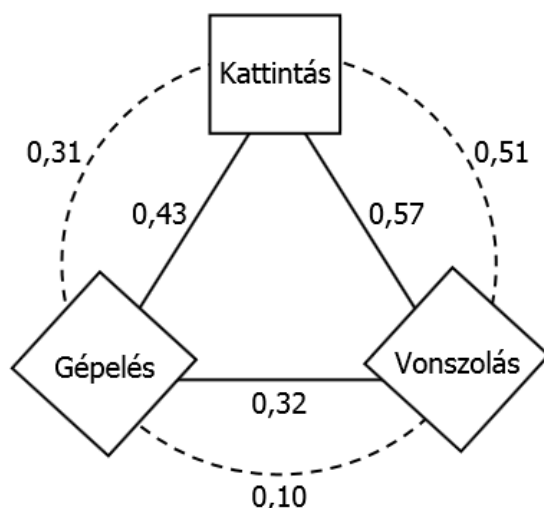
*Az egér- és billentyűzet-használati képességeket mérő feladatok személy-item térképe  
(minden 'x' 53 diákot reprezentál)*

Műveletenkénti bontásban megállapítható, hogy az első évfolyamos diákoknak átlagosan a legkisebb problémát a kizárólag kattintáson alapuló műveletek kivitelezése jelentette (átlag=67,7%, szórás=16,1; átlag\_nehézségi szint\_logit=-0,44), ezt követte (t=5,1, p<0,01) a gépelés, az írás művelete (átlag=66,1%, szórás=30,1; átlag\_nehézségi szint\_logit=0,19), majd a legnehezebben (t=8,1, p<0,01), de még mindig magas átlagteljesítménnyel a vonszolás művelete (átlag=62,0%, szórás=19,5, átlag\_nehézségi szint\_logit=0,25). A gépelés műveletére

vonatkozó eredmények értelmezése során szem előtt kell tartani, hogy e feladatok kizárólag néhány, nem különleges karakter begépelését kérték a diákoktól és nem több szó vagy mondatok írását.

Az egér- és billentyűzet-használati képességek fejlettségi szintjét a kattintás műveletét igénylő feladatokon mutatott teljesítmények jelezték előre leginkább ( $r=0,91$ ,  $p<0,01$ ), ezt követte a vonszolás ( $r=0,76$ ,  $p<0,01$ ), majd a gépelés műveletét kérő feladatokból ( $r=0,64$ ,  $p<0,01$ ) álló részteszt eredménye.

A kattintás és vonszolás műveletének sikeressége függött össze egymással leginkább, ami nem csökkent jelentősen az írás műveletére történő kontrollálás után sem. Ezzel szemben az írás és a vonszolás művelete közötti közepes összefüggés jelentős mértékben csökkent a kattintás műveletére történő kontrollálást követően (3.6. ábra). Mindezen eredmények, feltételezésünk szerint, másképp alakulnak táblagépek használata során, azonban, amíg az iskolákban a legelterjedtebb infrastruktúra az asztali számítógép, addig szükséges és releváns a diákok egér- és billentyűzet-használati képességszintjéhez történő igazodás, illetve szükség esetén azok fejlesztése.



3.6. ábra

*A kattintás, a vonszolás és a gépelés műveletei sikerességének kapcsolata első évfolyamon (korrelációs és parciális korrelációs értékek, az ábrán szereplő minden együttható  $p<0,01$  szinten szignifikáns)*

A műveletek közötti különbségek nagyságánál is jelentősebbnek bizonyult a műveleten belüli sikerességben lévő eltérés mértéke, ha külön kezeltük ugyanazon művelet időkorlátos belüli, illetve időkorlát nélküli megvalósításának sikerességét. Annak ellenére, hogy a két részteszt közötti együttjárás jelentős ( $r=0,56$ ,  $p<0,01$ ), mégis ugyanazon diákok időkorlát nélküli (átlag=74,8%,  $sd=16,0$ ; átlag\_nehézségi\_szint\_logit=-0,35) és időkorlátos (átlag=51,2%,  $sd=21,7$ ; átlag\_nehézségi\_szint\_logit=0,85) teljesítménye között minden művelet tekintetében jelentős mértékű volt az eltérés ( $t=89,9$ ,  $p<0,01$ ; műveletenkénti bontásban l. a 3.3. táblázatot). Átlagosan szignifikánsan könnyebb volt ugyanazon művelet időkorlát nélküli elvégzése, mint időkorlát alatt történő kivitelezése.

A 3.3. táblázat eredményei alapján a nagy elemre vagy elemekre való időkorlát nélküli kattintás, valamint egy kicsi űrlapelemre való kattintás egyáltalán nem jelent problémát a

diákoknak. Az ezen műveleteket tartalmazó feladatokon nyújtott teljesítményt nem befolyásolja jelentős mértékben egérhasználati képességük fejlettségi szintje, azaz bátran alkalmazhatóak azok a feladatok, ahol a diákoknak a válaszadás során nagy képre, képekre kell kattintani, illetve kattintás segítségével a színt kijelölni és ismételt kattintással képet színezní (pl. 3.7. ábra) vagy kattintás segítségével különböző feladatelemeket összekötni.

*3.3. táblázat. A kattintás, a vonzolás és a gépelés műveleteinek időkorláttal és időkorlát nélkül történő kivitelezési sikeressége első évfolyamon*

Művelet típusa			Időkorlát megléte	Átlag	Szórás	t (p)
Kattintás	nagy	egy	időkorlátos	79,5	40,4	-21,2
			nincs időkorlát	92,3	26,6	p<0,01
		több	időkorlátos	47,6	27,7	-92,3
			nincs időkorlát (pl. 1. ábra)	85,8	18,6	p<0,01
	kicsi (rádió- gomb, jelölő- négyzet)	egy	időkorlátos	79,3	40,5	-19,7
			nincs időkorlát	90,7	19,8	p<0,01
		több	időkorlátos (pl. 2. ábra)	42,5	29,3	43,5
			nincs időkorlát	62,6	26,9	p<0,01
Húzd és ejtsd (drag- and-drop)	nagy	egy	időkorlátos	70,2	34,1	-33,7
			nincs időkorlát	87,2	22,6	p<0,01
		több	időkorlátos	66,4	47,2	14,9
			nincs időkorlát	55,1	32,4	p<0,01
	kicsi	egy	időkorlátos	-	-	-
			nincs időkorlát	-	-	-
		több	időkorlátos	40,4	31,5	2,5
			nincs időkorlát (pl. 3. ábra)	38,7	48,7	p<0,05
Írás, gépelés	1 karakter			71,9	35,8	23,1
	több karakter (pl. 3.4. ábra)			62,4	32,8	p<0,01

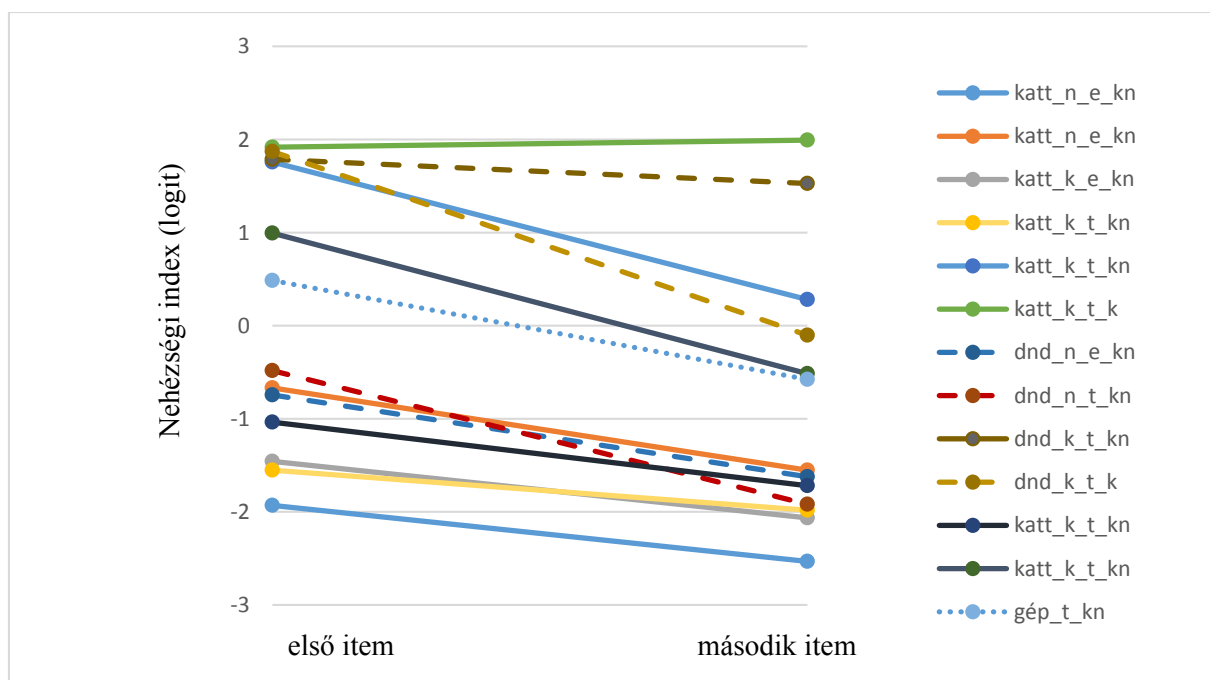
A vonzolás műveletét már körültekintőbben érdemes alkalmazni a legfiatalabb diákoknak szánt feladatok összeállítása során. Egy nagy elem mozgatása nem jelentett problémát számukra, azonban átlagteljesítményüket jelentős mértékben csökkentették a sok kicsi elem kicsi célpontba való mozgatását igénylő feladatok, melyek nélkül a vonzolásos részteszten elért teljesítmény 77% feletti. Ez azt jelenti, hogy más tartalom mérése során is teljesítménybefolyásoló lehet az egérhasználat fejlettségi szintje, ha a feladat sok kicsi elem kis célpontra való mozgatását kéri. Érdemes ezeket első évfolyamos diákoknak szánt tesztekben kerülni (pl. 3.3. ábra feladata).

A gépeléses feladatok esetében jobbak azok a feladatok, amelyek csak néhány karakter, esetleg egy egyszerű szó időkorlát nélküli beírását kívánják meg, mert az ennél bonyolultabb gépelési művelet már komoly teljesítménybefolyásoló erővel bírhat.

A műveletek feltételezett könnyű taníthatóságát támasztja alá az az eredmény, hogy összevetve az azonos műveletek elvégzését kívánó, tesztben korábban és később szereplő itemek nehézségi indexét, egyértelmű tendencia bontakozik ki (3.8. ábra): művelettől függetlenül a tesztben korábban szereplő, azonos művelet elvégzését igénylő item nehezebbnek bizonyult, mind a később szereplő item.



3.7. ábra  
Mintafeladat kattintással történő színezésre



3.8. ábra  
A kattintás, a vonszolás és a gépelés műveletei nehézségi indexének változása a teszt első és azonos műveletet kívánó második itemén (jelmagyarázat rövidítései: katt: kattintás, dnd: 'drag-and-drop', gép: gépelés, k: kicsi, n: nagy, e: egy elem, t: több elem: k: időkorlátos, kn: időkorlát nélküli)



A több kicsi vagy nagy alakzat vonzolását kívánó művelet esetén, amivel – nagy valószínűség szerint – először találkozott a kicsit nagyobb számítógépes tapasztalattal rendelkező diák is, átlagosan alacsonyabban teljesítettek az első évfolyamos tanulók a tesztben korábban elhelyezett feladatokon, mint a későbbiek.

A diákok átlagos teljesítménye első évfolyamon volt a legalacsonyabb, negyedikben a legmagasabb. Második és harmadik évfolyam között nem történt jelentős változás a diákok egér és billentyűzethasználati képességeinek fejlettségi szintje tekintetében (3.4. táblázat).

3.4. táblázat. A diákok egér és billentyűzethasználati képességszintjének változása évfolyamonkénti bontásban (mle)

Évfolyam	Átlag	t (évf.)	Min.	Max.	SD
1.	0,24	{1}<{2,3}<{4}	-2,90	3,04	0,93
2.	0,63		-2,04	3,70	0,85
3.	0,62		-2,18	2,77	0,88
4.	1,09		-2,18	5,77	1,21

Az évfolyamokon belül jelentkező teljesítménybeli különbségek – hasonlóan más képességterületek eredményeihez – jelentősebbek voltak, mint az évfolyamok átlagos teljesítményében történő változás mértéke. Mind a négy évfolyam kapcsán elmondható, hogy az adott évfolyam alacsonyan és magasan teljesítő diákjai messze az első évfolyamos átlagos teljesítmény alatt és a negyedik évfolyamos átlagos teljesítmény felett teljesítettek.

Nemek szerinti bontásban sem az egész minta vonatkozásában, sem évfolyamonkénti bontásban nem volt szignifikáns különbség a fiúk és lányok teljesítményében. Háttérváltozók közül az informatika jegy egyik évfolyamon sem bizonyult előrejelző tényezőnek a teszten nyújtott teljesítmény tekintetében. Az otthoni TV-k száma nem, a számítógépek száma egyedül második évfolyamon ( $r=0,27$ ,  $p<0,01$ ), míg a telefonok száma az első három évfolyamon ( $r=0,13$ ,  $0,14$  és  $0,31$ ;  $p<0,05$ ) volt meghatározó tényező. A diákok internet-hozzáférése minden egyes évfolyamon (1-4. évfolyam) teljesítménybefolyásoló erővel bírt ( $r=0,21$ ,  $0,25$ ,  $0,26$ ,  $0,24$ ,  $p<0,01$ ), azonban a számítógépezés és internetezés gyakorisága már egyik évfolyamon sem befolyásolta a teljesítményeket.

Összességében megállapítható, hogy már az első évfolyamos diákok körében is alkalmazhatóak a számítógép-alapú tesztek és a számítógép-alapú fejlesztőprogramok még akkor is, ha a diákoknak nem áll rendelkezésére táblagép, hanem az iskolákban elterjedtebb asztali számítógépeken dolgoznak is a teszten, illetve fejlesztő játékokon (Pásztor, 2014). A kutatás eredményei ugyanakkor felhívták a figyelmet arra, hogy esetükben a teszten nyújtott teljesítményüket nagyobb mértékben befolyásolja egér- és billentyűzet-használati képességeik fejlettségi szintje, ha a feladatok időkorlátosak, ezért javasolt ezen életkorban az időkorlátos feladatok kerülése. A válaszadás során az egy vagy több kattintást, vonzolást vagy néhány gyakori, egy billentyűzet leütésével elérhető karakter, esetleg egy szó beírását kívánó feladatok nem jelentenek problémát ebben a korban, míg kerülendő a hosszabb szavak, mondatok, illetve billentyűzetkombináció segítségével elérhető, általában ritka karakterek bevitelének kérése. Például, ha egy, a helyesírás szabályainak megfelelően nagybetűvel kezdődő nevet kell válaszul begépelniük, javasolt annak kisbetűvel történő begépelés esetén való elfogadása is (miután a nagybetű leütése már billentyűkombináció alkalmazását kívánja meg).

Az alsóbb évfolyamon tapasztalt teljesítménybeli növekedés egyik oka lehet, hogy a mintában résztvevő elsős diákok több mint felének a jelen tesztelés volt az első számítógép-alapú tesztelése, azaz semmilyen korábbi hasonló tapasztalattal nem rendelkeztek még. Ez a számítógép-alapú tesztelés elterjedésével várhatóan jelentős mértékben fog változni és a diákok vonatkozó technológiai műveltsége már az iskolába lépés után is jelentősebb mértékben fejlődni – akár az online tesztelés alkalmazásának mellékhatásaként is.

A pilot kutatás eredménye alapján egy évfolyamon belül tapasztalt képességszintbeli különbség egybecseng számos más képesség fejlődése során tapasztalt több évnyi fejlődésnek megfelelő mértékű különbséggel (*Csapó és Molnár, 2012*). Ennek hatására a legalacsonyabban teljesítő diákok képességszintje mind a négy vizsgált évfolyamon közel azonos, míg az átlagos és magasan teljesítők átlagos képességszintje évről évre nő. Miután a számítógép-alapú tesztek tömeges alkalmazása esetén a technológiai műveltség tekintetében alacsony képességszintűek más képesség tesztelése kapcsán is háttérbe kerülhetnek, ezért kiemelt feladat a technológiai műveltség és a technológiai műveltség vizsgált szeletének tekintetében a lemaradók felzárkóztatása.

A háttérváltozókkal történt összefüggés-vizsgálatok eredménye alapján megállapítható, hogy sem a diákok neme, sem általánosságban a szülők által biztosított gazdasági háttér nem befolyásolta technológiai műveltségük e szeletének fejlettségi szintjét. A háttérváltozók között kivételt képezett a telefonok száma, ami viszont feltételezhetően a család méretével, a testvérek számával, esetlegesen a saját telefon birtoklásával van összefüggésben. Ez természetesen további kutatás tárgyát is képezi. Feltételezésünkkel ellentétben a diákok géphasználati gyakorisága nem bizonyult meghatározó tényezőnek, csak az, hogy valamilyen gyakorisággal használnak-e számítógépet, internetet. Ez alapján tendenciaszerűen nem volt kimutatható az akár napi több órát számítógépezők előnye. Ennek oka lehet, hogy egyrészt a technológiai műveltség online tesztek megoldásához szükséges képességszintjét jóval kevesebb számítógép-használattal is el lehet érni, annak nem feltétele az akár napi több óra internetezés vagy számítógépezés, másrészt az érintett fiatal korban a napi több órát internetezők a táblagépek elterjedtsége miatt feltételezhetően érintőképernyős számítógépekkel játszanak és nem eger-, vagy billentyűzethasználatot igénylő eszközökkel.

E kutatás eredménye alátámasztotta azon feltételezésünket, hogy a ma iskolába lépő diákoknak általánosságban nem jelent problémát az eger és a billentyűzet használata. Azon diákoknak, akik korábban még nem vagy csak ritkán találkoztak számítógéppel és használták az egeret és billentyűzetet, néhány, a teszt előtt lévő példafeladat segítségével gyorsan és hatékonyan fejleszthető e képességük.

#### 3.1.4. Az eredmények értékelése, további kutatási feladatok

A teszt- és feladatbank szintű jószágmutatók, valamint a feladatok viselkedését jellemző indexek megerősítették azt az előzetes feltételezésünket, miszerint egyrészt kisiskolás diákok részére is kidolgozhatóak és alkalmazhatóak különböző tudás- és képességszintjük feltérképezésére alkalmas számítógép-alapú mérőeszközök, másrészt eger- és billentyűzet-használati képességszintjük tekintetében már a kisiskolás diákok is felkészültek az online tesztekkel történő empirikus kutatásokra. Az eger- és billentyűzet-használati képességeket nem tekinthetjük egy egydimenziós konstruktumnak, a kattintás, a vonszolás és a gépelés műveletei más-más készségek működtetését igénylik, ezért egyik működéséből nem feltétlenül

következtethetünk a másik hasonló szintű működésére, azaz műveleti szinten mindhárom mérése-értékelése, fejlesztése lényeges feladat.

A legfiatalabb tanulók részére kiközvetített tesztek kidolgozása során javasolt a kutatás eredményeinek figyelembe vétele: itemtípusok tekintetében elsősorban a kattintásra alapozó feladatok, esetlegesen a néhány gyakori karakter begépelését igénylő feladattípusok, illetve a nagy objektumok nagy területre történő mozgatását kívánó feladatok használata ajánlott. A kis elemek kis területre történő mozgatására épülő vonszolás típusú műveleteket, illetve a bonyolultabb gépelést igénylő feladatok vagy időkorlátos feladatok tesztbe történő integrációja esetén a teszten nyújtott teljesítményt erősebben befolyásolhatja a diákok egér- és billentyűzet-használati képességei, ezért javasolt azok kerülése az első évfolyamos diákoknak írt feladatokban. A diákok olvasási képességének fejlettségi szintjében lévő különbségek fülhallgató és a feladatok meghallgathatóságának biztosításával kiküszöbölhető.

### 3.2. Az iskolakezdés kulcsfontosságú előkészítéseinek online mérése kisiskolás korban

A felkészült iskolakezdés alapvetően meghatározza a későbbi évek tanulási eredményeit, a sikeres iskolai tanulmányok kulcsfontosságú feltétele (Nagy, 2000, 2007). Ezért világszerte jelentős erőfeszítések folynak annak érdekében, hogy minden gyermek felkészülten, azaz például az írás, olvasás, számolás elsajátításához szükséges előkészítésekkel megfelelő szinten rendelkezve kezdje meg az iskolát. E feltétel meglétének mérése, diagnosztizálása kutatási-fejlesztési feladat. Olyan megbízható, érvényes és könnyen alkalmazható tesztek (Snow, 2006), mérőeszközök kidolgozását igényli, amelyekkel pontosan meg lehet állapítani, hol tart egy gyermek a sikeres iskolakezdéshez szükséges készségek fejlődése tekintetében (Snow és Van Hemel, 2008).

A Szegedi Műhelyben négy évtizedes hagyománya van az iskolakészültség tesztek fejlesztésének, amelyek eddig kizárólag szemtől szembeni, azaz egyéni adatfelvételben valósultak meg. A továbblépés természetes lépése a technológia mérés-értékelésben nyújtott előnyeinek kiaknázása. Az online iskolakészültség tesztek fejlesztése során az idősebbek számára készült tesztekhez képest újabb problémákat is meg kell oldani: óvodások, írni-olvasni még nem tudó kisiskolások, korlátozott számítógépes tapasztalattal rendelkező gyermek számára kell tesztek készíteni olyan készségterületeken, ahol gyors és hirtelen minőségi változások következhetnek be. Továbbá bár az utóbbi évtizedben számos kutatás irányult a tesztelés közvetítő eszközének teljesítménybefolyásoló hatása vizsgálatára (l. 2.6. részt), azok főképp idősebb diákok papíralapú és számítógépes teszteken nyújtott teljesítménykülönbségeit térképezték fel, kevés tanulmány foglalkozik a nagyon fiatal tanulók technológiaalapú környezetben való tesztelésével (Choi és Tinkler, 2002; Carson, Gillon és Boustead, 2011; Fáyné, Hódi és Kiss, 2016) és kevés kutatási előzménye van a személyes megítélésen alapuló egyéni adatfelvételt igénylő tesztek és az online tesztek összehasonlításának is.

A fejezet középpontjában álló iskolaérettségi tesztet, a DIFER-t úgy tervezték, hogy a használata ne igényeljen különleges szakértelmet. Ez egyrészt a teszt sorozat hátránya is, ugyanis könnyen sérülhet a tesztek felvételének objektivitása. Előfordulhat, hogy a tesztet felvevő pedagógus kicsit másképpen olvassa fel az instrukciókat az egyes tanulóknak, és a válaszok pontozása is kismértékben változó lehet. A teszt sorozat alkalmazhatóságának másik korlátja, hogy csak személyesen és egyénileg végezhető el felvétele, így egy tanuló teljes körű felmérése akár több órát is igénybe vehet. A feladatok online közvetítésével, mindkét probléma

kiküszöbölhetővé válhat, előbbi esetén például előre rögzített hanginstrukciókkal (a szövegeket képzett szakemberek olvassák fel) és a válaszok automatikus pontozásával, utóbbi pedig a számítógép-alapú adatfelvétel adta csoportos tesztfelvételi lehetőséggel. A DIFER által mért hét területre vonatkozó tesztek közül kettő (szociális készségek és írásmozgás-koordináció) a jelenleg elterjedt technika alkalmazásával nem digitalizálható, egynek pedig csak részleges digitalizálása valószínűsíthető meg (elemi számlálás), ezért a fejezetben ismertetett elemzések négy teljes (beszédhanghallás, relációszőkincs, tapasztalati összefüggés-megértés, tapasztalati következtetés) és egy részteszt feladatai alapján összegyűjtött tapasztalatokra épülnek. A tesztekben előforduló hosszabb választ igénylő nyitott kérdéseket átalakítottuk többszörös választásos feladatokká az automatikus pontozás érdekében. A fejezetben ismertetett elemzésekbe bevontunk egy hatodik tesztet is (l. *Molnár, 2011b*), ami első évfolyamos diákok induktív gondolkodásának fejlettségi szintjét mérte és a teszt eredeti formájában nem szemtől szembeni, hanem papíralapú adatfelvétellel valósult meg.

A hat teszt különböző pszichológiai jellemzőket mér, és az online változataik különböző technológiai megoldásokat igényelnek. A mérőeszközöknek ez a változatos összetétele lehetőséget biztosít annak bemutatására, hogy a technológiaalapú tesztelés nem csupán kivitelezhető ilyen fiatal korban, hanem sok tekintetben még alkalmasabb is, mint a személyes felmérés vagy papíralapú tesztelés.

### 3.2.1. Célok, kutatási kérdések

A kutatások célja annak feltérképezése volt, hogy alkalmazhatóak-e ezen speciális, a területek sokfélesége miatt változatos, ezért számos kihívást is jelentő tesztek a hagyományos oktatási-nevelési gyakorlatba integrálva, az érintett, iskolába lépő célcsoportban csoportos adatfelvétel céljából. Milyen megbízhatósággal tehető ez, növelhető-e a tesztek reliabilitása a tesztíngerek feletti kontrol és a kiértékelés objektivitásának növelésével? Változik-e a tesztek validitása és a diákok eredményei azért, mert ők számítógép előtt és nem szemtől szemben, egyéni adatfelvétel során, vagy papíralapú környezetben oldják meg a tesztek? Ugyanúgy viselkednek-e az alacsonyabb és a magasabb képességszintű diákok a megváltozott környezetben? Mennyiben befolyásolja a tesztek tartalma és a teszteken belül alkalmazott itemtípusok a hagyományos és a számítógépes környezetben nyújtott teljesítményeket? Miután a kutatási kérdések megkívánták az eredmények hagyományos adatfelvételi eredményekkel való összevetését, ezért e kutatás a médiakutatások közé is sorolható, ahol nemzetközi szinten is hiánypótló. Kevés kutatás foglalkozik a személyes megítélésen alapuló egyéni adatfelvételt igénylő tesztek és teszteredmények valamint a számítógép-alapú tesztek, illetve számítógép-alapú teszteken nyújtott teljesítmények összehasonlításával.

### 3.2.2. Módszerek (minta, mérőeszköz, eljárások)

A szemtől szembeni és a számítógép-alapú adatfelvétel közötti hasonlóságokat és különbségeket górcső alá vevő kutatásban nyolc DIFER teszt (beszédhanghallás, a relációszőkincs négy változata, elemi számlálás, tapasztalati összefüggés-megértés és tapasztalati következtetés) digitalizálására, majd számítógép-alapú alkalmazására került sor. A mintát első osztályos (6-7 éves) diákok alkották. Az öt teszt digitális változatát más-más diákok oldották meg. A számítógépes adatfelvételben tesztenként kb. 400 diák vett részt (l. 3.5.

táblázat). A beszédhanghallás teszt esetében a személyes és a számítógépes adatfelvétel mintája megegyezett, a többi esetben nem. A relációszókinsz, az elemi számolás és a tapasztalati következtetés tesztek online eredményeit összevetettük egy egyéni adatfelvételen alapuló országos reprezentatív mintán felvett felmérés eredményeivel. A relációszókinsz teszt mind a négy változatát használtuk, ezért mindegyiket a teljes minta egynegyedén vettük fel.

### 3.5. táblázat. A DIFER szemtől szembeni és számítógépes adatfelvételének mintája

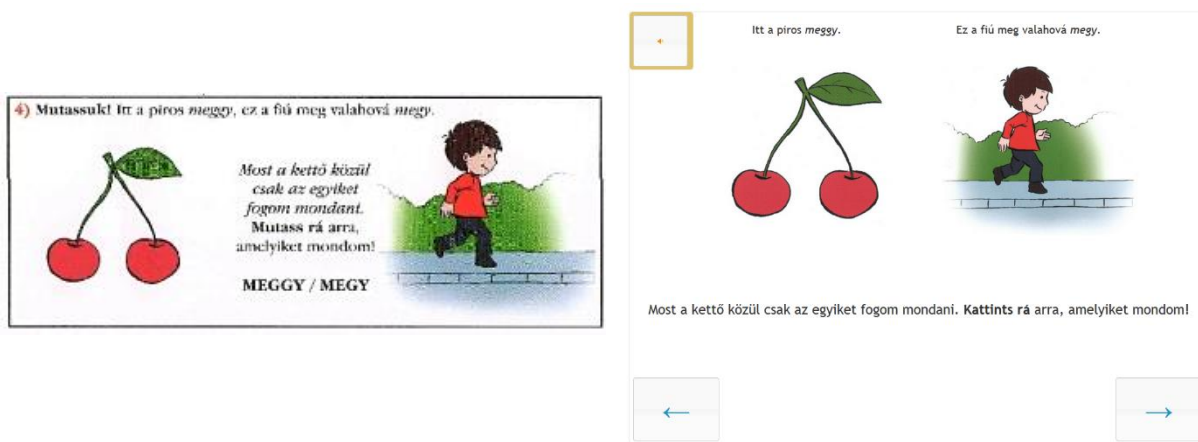
Teszt	N (egyéni/ PP)	N (számítógépes)
Beszédhanghallás	364	364
Relációszókinsz 1	1892	426
Relációszókinsz 2	1204	409
Relációszókinsz 3	1328	456
Relációszókinsz 4	999	449
Elemi számlálás	1895	435
Tapasztalati összefüggés-megértés	424	402
Tapasztalati következtetés	1094	416
Induktív gondolkodás	952	377

A kisiskolás diákok papíralapú, illetve számítógép-alapú tesztkörnyezetben történő viselkedésének leírására fókuszáló kutatásban egy korábban már többször alkalmazott (Molnár, 2006c, 2008c) és jól működő, kisiskolás diákok számára kidolgozott figurális tesztet digitalizáltunk. Mind a papíralapú, mind a számítógép-alapú adatfelvétel egy, az elemzések alapját képező mintánál nagyobb mintán valósult meg. Előbbi az MTA-SZTE Képességfejlődés Kutatócsoport longitudinális kutatássorozatának keretein belül (N=5156), utóbbi tágabb életkori intervallumban (N=2719). Az elemzésekhez háttérváltozók alapján egy, a számítógép-alapú adatfelvétel első évfolyamos részmintájához (N=377) illeszkedő, a tesztet papíralapon megoldott minta egy részmintájával dolgoztunk (N=952).

A tesztek számítógépesítése során különös figyelmet fordítottunk arra, hogy ne legyen szükség speciális számítógépes készségekre a teszt kitöltéséhez (l. 1. melléklet). A beszédhanghallás teszt 60 fonémakontraszt észlelést mérő itemet tartalmazott. A teszt első résztesztjében (22 item) a diákok képi illusztráció mellett egy mondatba ágyazva meghallgatták a feladatban szereplő szópárok (pl.: bont-pont) mindkét szavát, majd a két szó közül csak az egyiket hallották, és ki kellett választaniuk az elhangzott szót ábrázoló képet. A második részteszt (8 item) feladataiban az ejtés helyességéről kellett döntést hoznia a diákoknak, miközben a szó jelentését egy rajz ábrázolta. A hallott szópár egyike minden esetben helyes ejtésű szó, a másik helyes vagy torzított ejtésű szó volt (pl.: citom-citrom). Végül a harmadik és a negyedik részteszt valós (17 item) és álszavakat (13 item), szópárokat tartalmaztak képi segítség nélkül, melyek egymástól legfeljebb egy fonémában különböztek (pl.: rüg-rüf). A diákoknak el kellett döntenüik, hogy a szópárok két tagja azonos volt-e vagy sem.

A teszt hagyományos egyéni adatfelvétele során a tesztet felvevő személy olvasta fel a szavakat, mondatokat, a diákok pedig a szóhoz tartozó képre történő mutatással adták meg válaszukat, amit a tesztet felvevő személy értékelt, pontozott és feljegyzett a papíralapú válaszlapon (FF adatfelvétel – *face-to-face*). A számítógép-alapú változatban a gyerekek fejhallgatón keresztül hallgatták meg az instrukciókat, minden diák ugyanazt a hangot, egy képzett hangot hallotta (CB adatfelvétel – *computer-based*). A hangfájl lejátszása közben a

program piros keretet rajzolt a hagyományos tesztfelvételi módban mutatással jelzett kép köré. Ez, miután előre történő időzítéssel történt, mindenki számára egységes volt. A diákok válaszadása a megfelelő képre kattintással történt, aminek következtében a rendszer egy nagy piros pöttyöt hagyott a kiválasztott képen (l. 3.9. ábra).



3.9. ábra

*Mintafeladat a beszédhanghallás tesztből egyéni adatfelvétel és számítógép-alapú adatfelvétel esetén*

A relációszókinsz teszt mind a négy változata 24 itemet tartalmazott: teret (8 item; pl. belül, között), időt (4 item, pl. délután, tavasz) kifejező, mennyiségi (4 item; pl.: páratlan, kevés), hasonlósági (4 item; legrövidebb, azonos hosszúságú) és igekötői (4 item, pl.: rálép, belép) relációszókinsz fejlődésének leírására alkalmas feladatokat. A 3.10. ábrán látható egy feladat a számítógépes verzióból. A feladatok egy részének több mint egy helyes megoldása volt, amit a számítógépes program kezelni tudott.

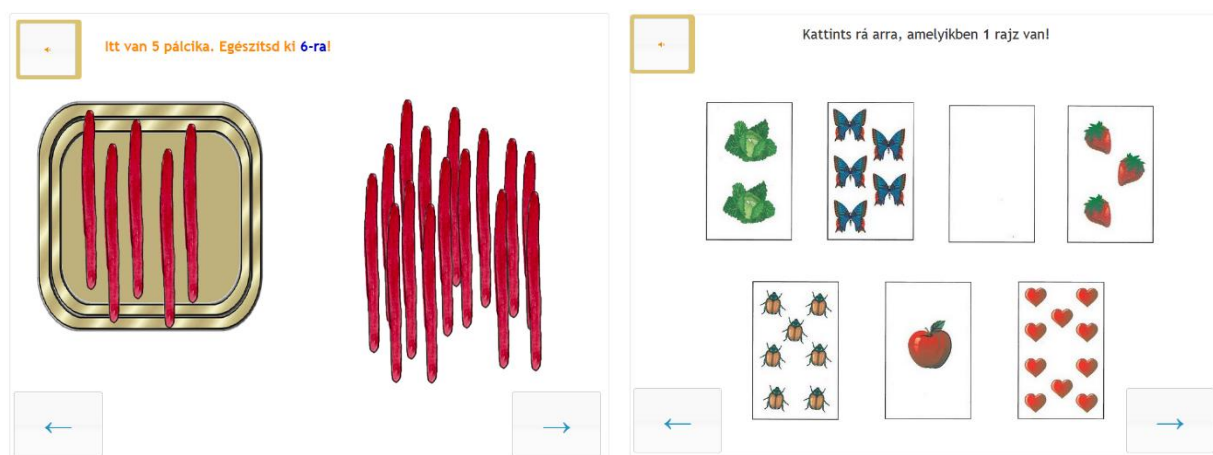
Az elemi számolási készség teszt hat résztesztből [21-ig számolás (egy item), 501-ig számolás (öt item), visszafelé számolás 501 és 0 között (nyolc item), húszas számkörben történő manipulatív számolás (tárgyakkal végzett műveletek; tizenegy item), tízes számkörbeli számképfelismerés (kilenc item), százas számkörbeli számolvasás (számok jelének felismerése; négy item)] hármát nem tudtunk a mért konstruktum megváltoztatása nélkül digitalizálni, miután az online platform még nem tudott szóbeli válaszokat kezelni. A teszt felvétele során a szóbeli választ igénylő feladatokat elhagytuk a számítógépes verzióban. A személyes és számítógépes adatgyűjtés menete megegyezett beszédhanghallás tesztrel leírtakkal. A 3.11. ábra illusztrálja a teszt számítógépes változatának egy-egy feladatát.

A tapasztalati összefüggés-megértést vizsgáló teszt eredeti, személyes adatfelvételre készült változata 32 kontextusba ágyazott, feleletalkotó feladatot tartalmazott. Ugyanezen képességtérületet mérő teszt számítógépes változata 32 többszörös választásos feladatból állt, lehetővé téve az azonnali kiértékelés megvalósítását. Az egyéni adatfelvétel során a tesztet felvevő egyszer, ismétlés nélkül felolvasta az instrukciókat, és a tanulóknak szóban kellett válaszolniuk a feltett kérdésre, amit az adatfelvételt végző személy pontozott és feljegyzett egy válaszlapon. Számítógépes módban a tanulók tetszés szerinti alkalommal meghallgathatták az instrukciókat és a lehetséges következtetéseket, mielőtt a megfelelő rádiógomba történő kattintással kijelölték válaszukat.



3.10. ábra

*Példafeladat a relációszőkincs teszt számítógépes változatából*



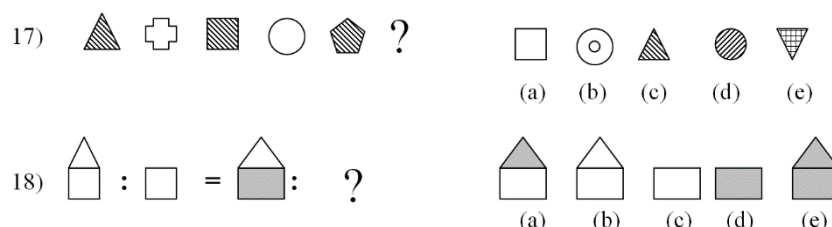
3.11. ábra

*Egy-egy példafeladat az elemi számolási készség tesztből*

A tapasztalati következtetés teszt 32 mindennapi szituációba ágyazott többszörös választásos feladata a legfontosabb kijelentés- és predikátumlogikai sémákat fedte le. Minden egyes feladat két premisszát és egy konklúziót tartalmazott. A diákoknak a két állítás alapján be kellett fejezniük a megkezdett harmadik állítást, azaz a konklúziót (l. Nagy, Józsa, Vidákovich és Fazekasné Fenyvesi, 2004a, 2004b).

A célzottan kisiskolás diákok részére kidolgozott induktív gondolkodás teszt 37 ítemet tartalmazott. A teszt kidolgozása során külön figyelmet fordítottunk annak nonverbális

jellegére, azaz minél több kép, ábra és figura használatára, míg az olvasnivaló szöveg mennyiségének minimalizálására (3.12. ábra). A teszt hat részesztből állt, melyek a *Klauer* (1989, 1991, 1993) által kidolgozott induktív gondolkodás meghatározására épültek: általánosítás, megkülönböztetés, többszemponú osztályozás, kapcsolatok felismerése, kapcsolatok megkülönböztetése és rendszeralkotás. (A teszt részletesebb ismertetését illetően lásd *Molnár*, 2006c, 2008c, 2009b).



3.12. ábra

*Mintafeladat az 1. évfolyamosok számára készült tesztből*

A tesztek számítógépre adaptálása során törekedtünk arra, hogy annak minél több tulajdonságát megőrizzük. Megtartottuk a teszt linearitását, azaz minden egyes diák ugyanabban a sorrendben kapta ugyanazokat a feladatokat, ugyanakkor az alkalmazott webes felület nem tette lehetővé, hogy az egyszerre a képernyőn látható feladatok mennyisége megegyezzen a papíralapú tesztfüzet egy oldalán található feladatokéval. A tesztfeladatok közötti navigálást mindkét esetben egy előre és egy visszafelé lépést lehetővé tevő navigációs gombbal valósítottuk meg. A digitalizálás során hasonló módon történt a feleletválasztó itemek átalakítása. A papíron karikázással megvalósítandó válaszadást a számítógépen képre, képekre való kattintással kellett megvalósítani és kiválasztatni a helyesnek ítélt megoldást.

A DIFER tesztek online adatfelvétele az *eDia* elektronikus diagnosztikus platform segítségével, míg az induktív gondolkodás teszt kiköszvetítése a TAO platform alkalmazásával (*Plichart, Jadoul, Vandenabeele és Latour*, 2004; *Farcot és Latour*, 2008) történt. Mindkét kutatás számítógép-alapú adatfelvétele interneten keresztül, az iskolák számítógépes termeiben zajlott. A számítógépek, operációs rendszerek és böngészők sokféleségének hatásai nem kerültek szisztematikus vizsgálat alá, de ez a körülmény nem okozott feltűnő problémát, és nem volt érezhető hatással az eredményekre. A tesztfelvétel 20-45 percet vett igénybe az adott tesztől függően. A válaszokat automatikusan pontozta a program, és a tanulók azonnali visszajelzést kaptak (helyes válaszok aránya) a teszt végén.

A DIFER kutatásban érintett tesztjeinek személyes egyéni felvétele tanulónként hozzávetőlegesen egy órát vett igénybe, a számítógépes változathoz tipikusan két iskolai tanóra volt szükség. Az induktív gondolkodás teszt megoldása mind papír, mind számítógépalapon egy teljes tanórát vett igénybe.

Különböző kutatási elrendezést alkalmaztunk a tesztek felvételekor, ezért más és más elemzéseket végezhattünk a különböző tesztekre vonatkozóan. A klasszikus tesztelméleti elemzések mellett alkalmaztunk valószínűségi tesztelméleti eljárásokat és strukturális egyenletek segítségével történő strukturális és mérési invariancia meglétét vizsgáló konfirmációs faktoranalíziseket is. Utóbbi előfeltétele volt a klasszikus és valószínűségi tesztelméleti elemzéseknek (a disszertáció későbbi részeiben is), miután az eredmények



összehasonlítása csak akkor lehetséges, ha a vizsgált konstruktum szerkezetét nem befolyásolja a közvetítő eszköz változása (Byrne és Stewart, 2006), azaz a csoportok között realizálódó esetleges különbségeket nem a tesztek megváltozott pszichometria jellemzői indukálják (Brown, 2006).

### 3.2.2.1. A mérési invariancia tesztelése – strukturális egyenletekkel történő elemzések

A mérési invariancia meglétének statisztikai ellenőrzésére számos megközelítés és eljárás érhető el (Schroeders és Wilhelm, 2011). A legújabb és legprominensebb módszerek közé sorolható a valószínűségi tesztelméleten belül a különböző itemműködést (DIF – *differential item-functioning*) vizsgáló elemzések (Raju, Laffitte és Byrne, 2002), illetve a strukturális egyenletekkel történő modellezés (SEM - structural equation modeling; Byrne, 2012) keretein belül például a többszoportos megerősítő faktoranalízis (MGCFA – *multi-group confirmatory factor analysis*; Steenkamp és Baumgartner, 1998; Vandenberg és Lance, 2000).

A mérési invariancia tesztelése alapvető fontosságú médiahatáskutatások során. Az eredmények összehasonlítása során ugyanis feltételeznünk kell, hogy a két adatfelvételi módon mért látens változó azonos felépítésű, struktúrájú és működésű (Byrne, 2008). Ha ezek a feltételek nem állnak fenn, vagy nem tudjuk, hogy fennállnak-e, nem lehetünk abban biztosak, hogy ugyanannak a konstruktnak a mérését valósítottuk meg papíralapú és számítógépes környezetben.

A jelen kutatásban a független szempont szerint csoportosított egyedek közötti (*between-subject design*) többszoportos megerősítő faktoranalízist, valamint az egyszapoportos időben eltolat megerősítő faktoranalízist alkalmaztunk.

Kategoriális változók esetében a mérési invariancia tesztelése (Vandenberg és Lance, 2000; Muthén és Muthén, 2010) több lépésben történik többszoportos megerősítő faktoranalízisek sorozata segítségével (3.6. táblázat). A mérési modell (*basic model*) meghatározása után három beágyazott modell egyre szigorúbb feltételek melletti ellenőrzésére (l. Byrne és Stewart, 2006) került sor (gyenge metrikus invariancia tesztelés kategoriális adatok esetén nem ajánlott, ezért arra nem került sor; Greiff, Wüstenberg, Molnár, Fisher, Funke és Csapó, 2013; Schroeders és Wilhelm, 2011).

Az invariancia első szintje a konfigurális invariancia, ami azt vizsgálja, hogy mérési modell struktúrája invariáns-e a csoportok között (Byrne, 2008), azaz jelen esetben mindkét csoportban azonos változókból és azonos módon épül-e fel, a diákok CB és FF/PP környezetben ugyanúgy konceptualizálják-e a mért konstruktnak (Milfont és Fischer, 2010), ugyanazt a fogalmi keretet használják-e a teszt itemeinek megválaszolása során (Wu, Li és Zumbo, 2007; Vandenberg és Lance, 2000), az egyes vizsgált változók ugyanazt a látens változót mérik-e mindkét csoportban, mindkét adatfelvételi környezetben. A konfigurális invariancia biztosítja, hogy a bemeneti változók és a modellünk struktúrája mindkét csoportban megegyezik (Brown, 2006). A konfigurális invariancia megléte esetén ugyanazon item CB és FF/PP változata ugyanazon látens változó indikátora mindkét adatfelvételi környezetben (faktorsúlyok esetén megengedett az eltérés), ezért a konfigurációs invariancia modellben a faktorsúlyok és az interceptek szabad becslése történik, miközben a reziduális varianciáértékeket minden csoportban 1-re, a faktor átlagokat 0-ra rögzítjük.

3.6. táblázat. Kategoriális változók esetén a mérési invariancia tesztelésének lépései

Kategoriális változók	Faktorsúly	Intercept	Reziduális variancia	Faktor átlagok
Konfigurális invariancia	szabadon közelített	szabadon közelített	1-re rögzített	0-ra rögzített
Metrikus invariancia	csoportok között azonosan tartott	csoportok között azonosan tartott	PP: 1-re rögzített CB: szabadon közelített	PP: 0-re rögzített CB: szabadon közelített
Szigorú metrikus invariancia	csoportok között azonosan tartott	csoportok között azonosan tartott	1-re rögzített	FF: 0-ra rögzített CB: szabadon közelített

Megj.: PP: papíralapú, CB: számítógép-alapú, FF: szemtől szembeni adatfelvétel

Az invariancia tesztelésének második szintjén, a metrikus invariancia vizsgálata során, a mért változók által meghatározott látens faktorok faktorsúlyainak összevetése valósul meg. Tesztelése csak akkor releváns, ha a konfigurális invariancia áll, azaz modellje illeszkedik az adatokra. Megléte azt mutatja, hogy a látens változók varianciája és a látens változók közötti kovariancia azonos-e a két csoportban, azaz a CB és az FF/PP modell esetén azonosak-e a faktorsúlyok és az interceptek (tengelymetszetek). Egy ilyen feltételrendszerrel közelített modell (a két modellben a látens változókból az indikátorokba tartó regressziós utak együtthatói azonosak) illeszkedése szignifikánsan rosszabb-e az adatokhoz, mint a konfigurális invariancia modellje volt. Ha a modell illeszkedése az egyenlővé tett regressziós együtthatók ellenére sem romlik a konfigurális állandóság modelljének illeszkedéséhez képest, akkor fennáll a metrikus invariancia (Byrne, 2010). A két csoportban a látens változók skálája azonos léptékű (Blunch, 2010) és irányú, azaz a csoportok között összehasonlíthatóvá válnak a látens változóra irányuló, vagy abból induló hatások, utak, regressziós együtthatók (Steinmetz, Schmidt, Tina-Booh, Wiecek és Schwartz, 2007). A metrikus modell tesztelése során a faktorsúlyokat egyenlő értéken tartottuk, a reziduális variancia értékeit a FF (vagy PP) csoportban 1-re rögzítettük, míg a CB csoportban szabadon közelítettük, a faktorátlagokat is az FF (vagy PP) csoportban 0-ra rögzítettük, a CB csoportban pedig szabadon közelítettük (Muthén és Muthén, 2010). [Folytonos változók esetén az invarianciavizsgálat második lépcsőfoka a gyenge faktoriális invariancia (*weak factorial invariance*) meglétének tesztelése lenne, azonban ennek alkalmazása kategoriális változók esetén nem ajánlott (Muthén és Muthén, 2010; Schroeders és Wilhelm, 2011).]

Ha a modifikációs indexek felülvizsgálata során megállapítható, hogy a modellilleszkedés jelentős mértékben javulna abban az esetben, ha néhány item interceptjét felszabadítanánk, azaz megengednénk, hogy azok különbözzenek a CB és az FF/ PP csoportban, akkor részleges metrikus invariancia meglétéről beszélhetünk. A metrikus invariancia megléte esetén hasonlíthatjuk csak össze a látens változók átlagait (Byrne és Stewart, 2006).

A szigorú metrikus invariancia, ami a csoportátlagok összehasonlíthatóságának feltétele, az invariancia harmadik szintjének tesztelése csak akkor releváns, ha fennáll a metrikus invariancia. A szigorú metrikus invariancia megléte biztosítja, hogy a látens változók skálájának nemcsak a mértéke egyezik meg (metrikus invariancia), hanem a skála viszonyítási pontja is azonos, azaz a látens változókat mérő két skála megegyezik egymással (Blunch, 2010). Ha a látens változó skálái megegyeznek, a látens változók átlagai összehasonlíthatóak

egymással (Byrne, 2008). Az invariancia e fokának tesztelése során a látens változó indikátoraihoz tartozó tengelymetszetek egyenlősége kerül górcső alá (Steinmetz és mtsai, 2007). A szigorú metrikus invariancia tesztelése esetén mind a faktorsúlyokat, mind az intercepteket egyenlő értékeken tartottuk a csoportok esetében, a reziduális variancia értékeket mindkét csoportban 1-re rögzítettük a faktor átlagokat az FF (vagy PP) csoportban 0-ra rögzítettük, a CB csoportban szabadon közelítettük. [A szigorú metrikus invariancia megléte médiahatáskutatások esetén nem előfeltétel, azaz nem feltétele az eredmények összehasonlíthatóságának (Meredith, 1993).]

A fejezetben, hasonlóan a disszertációban bemutatott többi kutatás eredményeinek elemzéséhez a klasszikus tesztelméleti elemzéseket az SPSS 21.0 szoftvercsomaggal, a valószínűségi tesztelméleti elemzéseket a ConQuest programmal, míg a strukturális egyenletekkel kapcsolatos (SEM) elemzésekhez az MPlus 6.0 programmal végeztük.

### 3.2.3. A DIFER tesztek online változatával végzett mérések eredményei

A tesztek belső konzisztenciája (Cronbach- $\alpha$ ) megfelelőnek bizonyult mind az egyéni, mind a számítógép-alapú adatfelvétel során. Előbbi esetben a reliabilitásmutatók értékei 0,71 és 0,89 között mozogtak, utóbbi esetben pedig 0,74 és 0,94 között. Általában az online adatfelvétel mutatói magasabbnak bizonyultak, mint a szemtől szemben egyénileg történt adatfelvétel eredményei alapján számolt mutatók értékei (3.7. táblázat).

3.7. táblázat. Az egyéni és a számítógép-alapú adatfelvétel eredményeinek megbízhatósági mutatói

Teszt	Itemek száma	Cronbach- $\alpha$	
		Egyéni adatfelvétel	Számítógépes adatfelvétel
Beszédhanghallás	60	0,887	0,938
Relációszőkincs 1	24	0,796	0,844
Relációszőkincs 2	24	0,706	0,824
Relációszőkincs 3	24	0,749	0,809
Relációszőkincs 4	24	0,768	0,851
Tapasztalati összefüggés-megértés	32	0,743	0,831
Tapasztalati következtetés	32	0,885	0,904
Elemi számolás	13	0,812	0,739

Mind a hagyományos, mind a számítógép-alapú adatfelvételen a legmagasabb mutatóval a beszédhanghallás teszt rendelkezett, ahol a személyes adatfelvételen alapuló eleve magas megbízhatósági mutató (Cronbach- $\alpha$ =0,89) magasabbra emelkedett (Cronbach- $\alpha$ =0,94). A javulást a standardizált beszédhangingereknek tulajdoníthatjuk. Az egyéni adatfelvétel során viszonylag alacsony megbízhatósági értéket mértünk a relációszőkincs, a tapasztalati következtetés és a tapasztalati összefüggés-megértés teszteken, amelyek közül az első kettő jelentős, a harmadik pedig némi javulást mutatott a számítógépes változatban.

Az elemi számolási készség teszt esetében ezzel szemben a megbízhatósági mutató jelentős romlását figyelhetjük meg, miután a hagyományos verzió több feladatát is ki kellett

hagynunk a számítógépes változatból. Ennek ellenére a rövidített teszt egyéni felvétele is viszonylag magas megbízhatóságot mutatott (0,81), azaz az alacsony elemszám nem ad magyarázatot a számítógépes változat alacsonyabb megbízhatóságára.

Miután a beszédhanghallás tesztet ugyanazon minta mindkét adatfelvételi módban megoldotta, ezért az egycsoportos időben eltolt megerősítő faktoranalízist alkalmaztunk a mérési invariancia tesztelése során. Első lépésben külön-külön teszteltük konfirmációs faktoranalízis (CFA) alkalmazásával az egy-, a két-, a három- és a négydimenziós elméleti modelleket (3.8. táblázat), melyeket az adatok kategoriális jellege miatt egy speciális  $\chi^2$ -próbával hasonlítottunk össze (WLSMV közelítő eljárás mellett alkalmazandó speciális DIFFTEST segítségével; *Muthén és Muthén*, 2010).

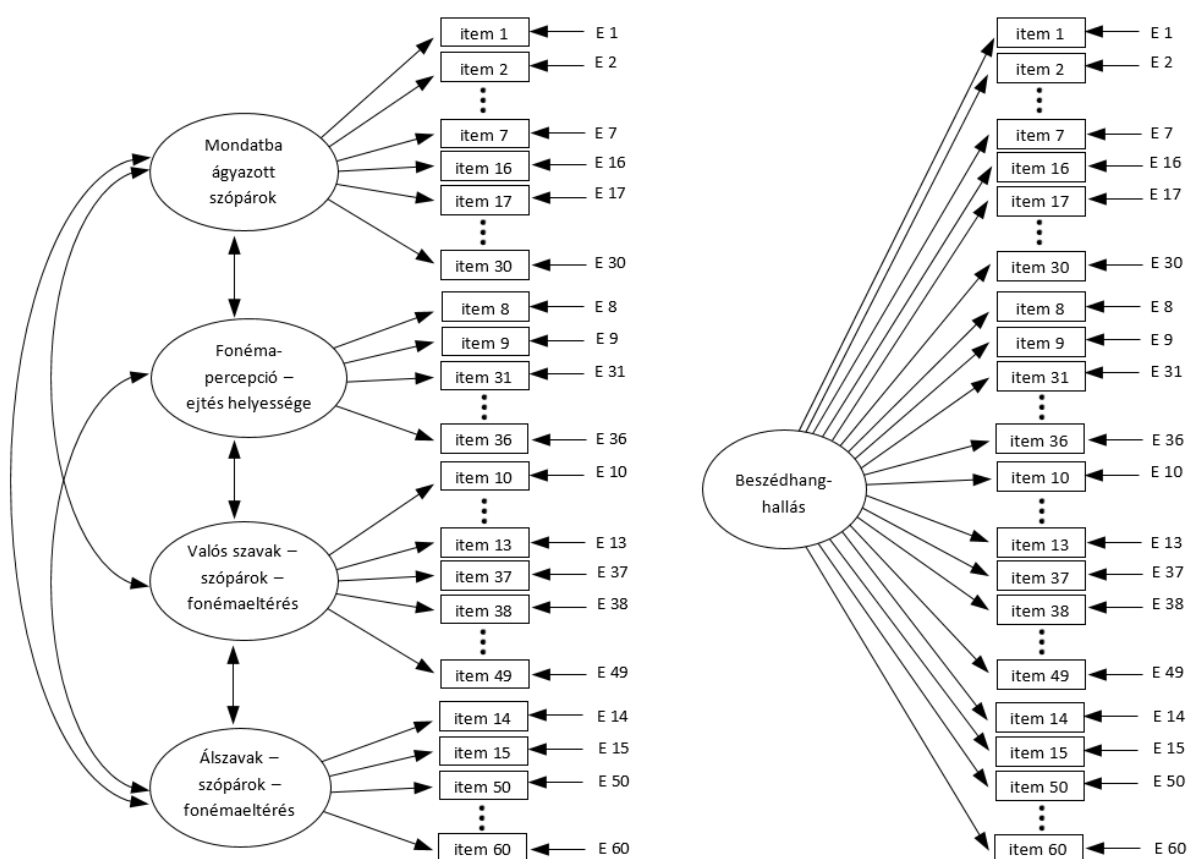
3.8. táblázat. A beszédhanghallás (BHH) teszt elméleti modelljeinek illeszkedésmutatói (számítógép-alapon)

Modell	$\chi^2$	df	p	CFI	TLI	RMSEA
Itemszinten						
1 dimenziós (Részteszt 1, 2, 3, 4)	21980,6	1127	0,001	0,901	0,897	0,034
2 dimenziós (Részteszt 1,2 – 3,4)	18170,3	1126	0,001	0,936	0,934	0,027
3 dimenziós (Részteszt 1 – 2 – 3,4)	16830,7	1124	0,001	0,948	0,946	0,025
4 dimenziós (Részteszt 1 – 2 – 3 – 4)	16650,2	1121	0,001	0,956	0,954	0,024
Részteszt szintjén						
BHH (dim 1 – dim 2 – dim 3 – dim 4)	60,205	2	0,001	0,996	0,988	0,054

Az eredmények értelmében a kétdimenziós modell szignifikánsan jobban illeszkedett az adatokhoz, mint az egydimenziós ( $\chi^2=45,5$ ;  $df=1$ ;  $p<0,001$ ), a háromdimenziós szignifikánsan jobban illeszkedett, mint a kétdimenziós ( $\chi^2=55,4$ ;  $df=2$ ;  $p<0,001$ ) és a négy dimenziós modell illeszkedése szignifikánsan erősebb volt, mint a háromdimenziósé ( $\chi^2=19,0$ ;  $df=3$ ;  $p<0,001$ ), ezért a médiumtól függetlenül (a négydimenziós modell illeszkedésmutatói FF alapon: RMSEA=0,016, CFI=0,946, TLI=0,944;  $\Delta CFI_{FF\_CB}=0,01$ ) item és részteszt szinten is jól illeszkedő négydimenziós modellel dolgoztunk a további elemzésekben. Az egy- és a négydimenziós modell felépítését mutatja a 3.13. ábra.

A teszt mérési invarianciájának tesztelése során a metrikus invariancia modell a konfigurális modellhez képest gyengébben illeszkedett, ugyanakkor a részleges metrikus invariancia modell nem mutatott romlást a modellilleszkedés vonatkozásában. A részleges szigorú metrikus modell ismételen gyengébb illeszkedésűnek bizonyult, azonban ennek teljesülése nem feltétele a teszteken nyújtott teljesítmények összevetésének.

A relációszókincs, elemi számolás és tapasztalati következtetés tesztekénél mind a metrikus, mind a Lagrange-multiplikátor teszt alapján felállított részleges metrikus invariancia modell szignifikánsan gyengébben illeszkedett (3.9. táblázat), mint a konfigurális modell. Nem beszélhetünk mérési invarianciáról e tesztek kapcsán.



3.13. ábra

*A beszédhanghallás teszt egy és négydimenziós elméleti faktorstruktúrája*

A két tesztelési módból adódó különbségeknek számos oka lehet. Annak ellenére, hogy szigorúan véve a mérési invarianciaelemzések eredménye alapján csak a beszédhanghallás teszten nyújtott teljesítmények fejezhetőek ki egy skálán, azaz hasonlíthatók össze, a tesztek tartalmának, a feladatok típusának, és a vizsgált személyek jellemzőinek hatásai elemzése céljából a többi teszten nyújtott teljesítményeket is görcső alá vettük. Empirikusan bizonyított validitása miatt kiemelten a beszédhanghallás teszt eredményeire fókuszáltunk. Az eltérő módokon felvett tesztek lehetővé tették a személyes és számítógépes változatok páronkénti összehasonlítását.

Elemeztük a tesztelés megváltozott módjának hatását a diákok teljesítményre (3.10. táblázat). Miután a DIFER kritériumorientált tesztek közül áll, ezért az első osztályos tanulók többsége már elérte az iskolakezdéshez szükséges fejlettségi szintet. Az eredmények alapján nyolc esetből hét esetben a számítógép-alapú teszteken gyengébb eredményt értek el a tanulók, mint a hagyományos, személyes adatfelvételen alapuló teszteken.

A beszédhanghallás teszt igen magas eredményeiben jelentős mértékű esés volt megfigyelhető (91,4 százalékról 82,6 százalékos teljesítményre). A relációsózikincs teszt négy változatában mindkét irányú változásnak tanúi lehettünk: jelentős teljesítménycsökkenés két esetben, kismértékű növekedés a másik kettőben. Az elemi számolás esetében nem észleltünk számottevő változást (3.10. táblázat).

3.9. táblázat. A tesztek mérési invarianciáját jellemző modellek illeszkedésmutatói

Teszt	Modell	$\chi^2$	df	CFI	TLI	RMSEA	$\Delta\chi^2$ *	$\Delta df$ *	p
Beszédhang-hallás	(1)	29050,7	2628	0,908	0,906	0,017	-	-	-
	(2)	29500,2	2663	0,905	0,904	0,017	60	35	<0,05
	(2.1)	29350,9	2661	0,909	0,908	0,017	440,8	33	>0,05
	(3)	32310,4	2700	0,824	0,824	0,024	287	72	<0,05
Relációsózkinc	(1)	3890,9	97	0,971	0,961	0,051	-	-	-
	(2)	4900,5	108	0,962	0,954	0,055	890,2	11	<0,01
	(2.1)	3960,0	105	0,971	0,964	0,049	270,4	8	<0,01
	(3)	10270,1	119	0,910	0,900	0,081	4630,9	22	<0,01
Elemi számolás	(1)	460,2	17	0,996	0,997	0,038	-	-	-
	(2)	2430,4	22	0,984	0,978	0,093	1280,2	5	<0,01
	(2.1)	1570,0	20	0,990	0,985	0,076	700,8	3	<0,01
	(3)	2680,9	27	0,983	0,981	0,087	1690,6	10	<0,01
Tapasztalati következtetés	(1)	1850,8	136	0,980	0,973	0,030	-	-	-
	(2)	2930,1	146	0,942	0,927	0,049	750,9	10	<0,01
	(2.1)	2640,4	142	0,951	0,938	0,046	560,4	6	<0,01
	(3)	3480,1	166	0,928	0,921	0,051	34	14	<0,01

Modell: (1) Konfigurális invariancia, (2) Metrikus invariancia, (2.1) Részleges metrikus invariancia, (3) Szigorú metrikus invariancia. \*  $\Delta\chi^2$  és  $\Delta df$  értékeket az MPlus DIFFTEST eljárásával számoltuk.

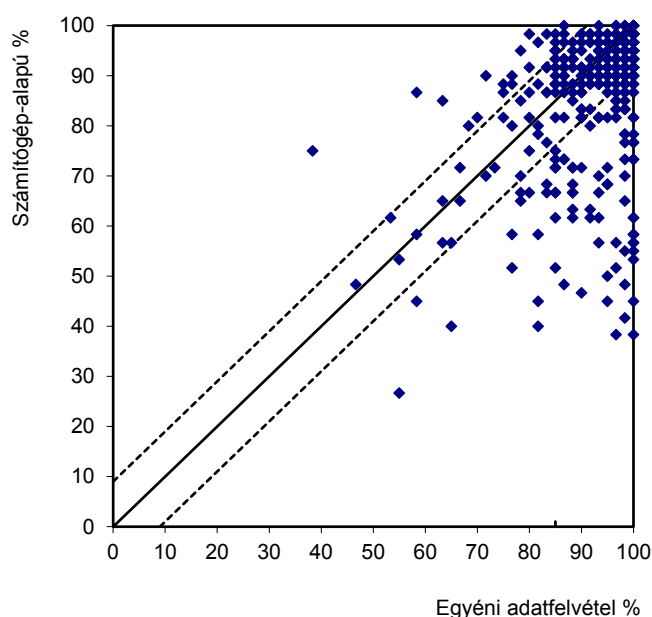
3.10. táblázat. Az egyéni és számítógépes adatfelvétel eredményei közötti különbségek

Teszt	Egyéni adatfelvétel (%)		Számítógép-alapú adatfelvétel (%)		d	t	p
	Átlag	Szórás	Átlag	Szórás			
Beszédhanghallás	91,35	9,88	82,61	16,80	0,59	-8,79	0,001
Relációsózkincs 1	80,86	15,44	78,03	17,86	0,17	-3,32	0,01
Relációsózkincs 2	83,88	12,06	73,06	18,23	0,72	-13,62	0,001
Relációsózkincs 3	84,07	13,39	72,86	18,16	0,71	-13,98	0,001
Relációsózkincs 4	84,84	13,62	81,55	17,59	0,21	-3,86	0,001
Tap. összefügg.-m.	70,69	14,34	63,59	20,87	0,39	-5,72	0,001
Tapasztalati köv.	69,64	20,66	55,93	23,64	0,58	-8,61	0,001
Elemi számolás	87,35	18,46	88,90	14,84	-0,09	1,52	n.s.

A különbségnek két magyarázata is lehet. Egyrészt a tesztet felvevő pedagógusok gyakran magasabb pontszámokat adnak a gyerekeknek, mint a számítógép automatikusan (és objektíven) kiszámolt pontszámai, vagy a tanulóknak nehézséget jelentett a számítógép

kezelése, ami rontott teljesítményükön. Miután sem a következő részben ismertetett, azonos korosztály bevonásával végzett kutatásban, sem az elemi számolás (magas pontszám mindkét verzióban) esetén nem találtunk utóbbira utaló jelet, ezért a számítógép kezelési nehézség nem lehet kielégítő magyarázat a teljesítménykülönbségekre. Valószínűbb, hogy a teszteket felvevő pedagógusok elnézőbbek voltak a gyerekek válaszaival szemben.

A közvetítő eszköz teljesítménybefolyásoló hatásainak alaposabb feltérképezése érdekében megvizsgáltuk, hogy detektálható volt-e valamilyen irányú szabályszerűség a különböző képesség-fejlettségi szinteken lévő diákok esetében. A beszédhanghallás teszt kutatási elrendezése lehetőséget adott az eredmények tanulónkénti összehasonlítására. A tanulók eredményeit egymás függvényében a 3.14. ábra szemlélteti. A vízszintes tengely a személyes módban, a függőleges tengely pedig a számítógépes módban elért eredményeket mutatja.



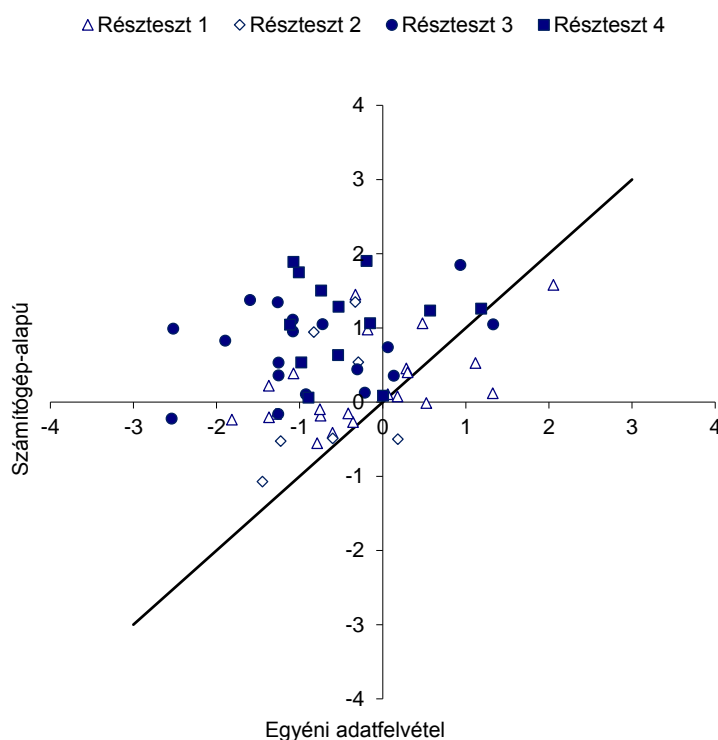
3.14. ábra

*A tanulók hagyományos illetve számítógépes módban nyújtott teljesítményei a beszédhanghallás teszten*

A tanulók hagyományos és számítógépes környezetben mutatott teljesítménye között szignifikáns, de viszonylag alacsony volt a korrelációs együttható ( $r=0,35$   $p<0,001$ ). Sok esetben a tanulók hagyományos tesztelési módban elért teljesítménye 100 százalékos, vagy 90 százalék fölötti volt, míg ugyanezen diákok számítógépes környezetben nyújtott teljesítménye messze elmaradt ettől a szinttől. Ez arra utal, hogy az adatfelvételt végző pedagógusok feltételezhetően hajlamosak voltak akkor is elfogadni a tanulók választ, amikor a gyerekek nem a helyes választ adták meg, vagy a kérdés feltevésekor egyéb eszközökkel segítettek a diákoknak: például az instrukciók felolvasásának módjával, a helyes és a helytelen válaszlehetőség artikulálásával (szemben a semleges előre rögzített hanggal), arckifejezéssel, egyéb mozdulatokkal, testbeszéddel. A tapasztalt jelenség másik oldalról történő megközelítése, hogy az emberi tesztvezető megítélése a pontosabb, és a számítógép – többek

között a számítógépes képességeik miatt – alulbecsüli teljesítményüket ellentmond a korábbi kutatási eredményeknek.

Az itemszintű különbségekre irányuló elemzések eredményeit mutatja a beszédhanghallás teszt kapcsán a 3.15. ábra. Az item abszcissa koordinátája a hagyományos adatfelvétel alapján számolt nehézségi indexe, ordináta koordinátája a számítógép-alapú adatfelvételen nyújtott teljesítmény alapján számolt indexe. Azok a feladatok, amelyek könnyebbnek bizonyultak az egyéni adatfelvétel során az ábrán a folytonos egyenessel jelzett vonal fölött helyezkednek el.



3.15. ábra

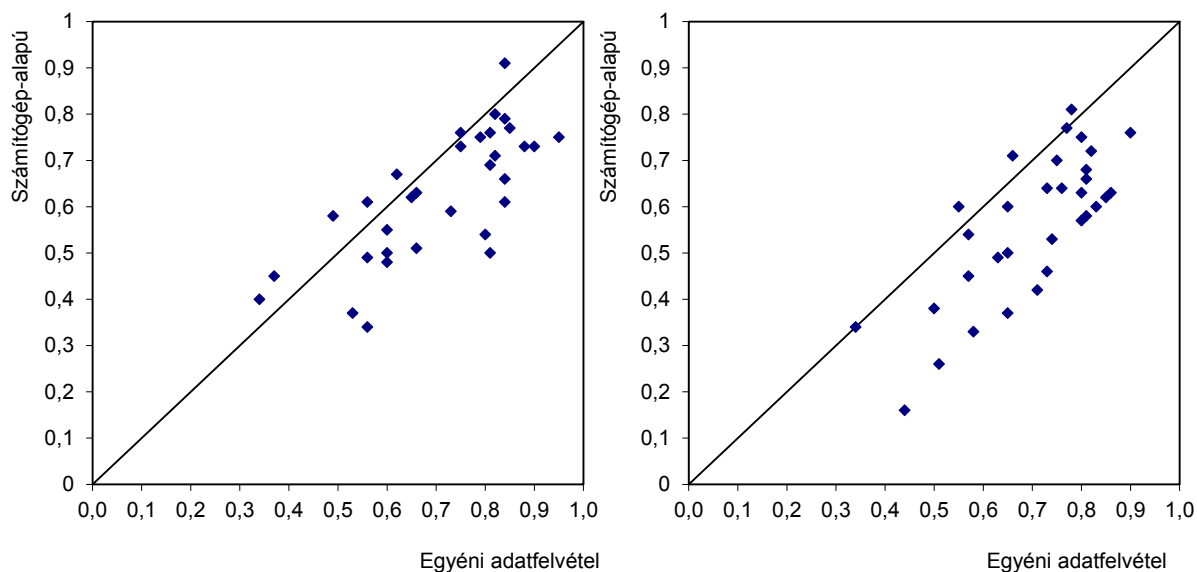
*Az IRT modellel számított itemparaméterek összehasonlítása a beszédhanghallás teszt egyéni és számítógépes adatfelvételi módjában*

A legnagyobb eltérések az álszavakra építő (4. részteszt) feladatok esetén voltak detektálhatóak. Ugyanez az effektus volt megfigyelhető az olyan szavak esetében, ahol a kontextus hiányát részben helyettesíthette a tanító jószándékú viselkedése, míg a teljes mondatokban szereplő szavaknál a személyes tesztelés hatása kevésbé volt nyilvánvaló. (Csak az utóbbi esetben volt szignifikáns a korreláció az itemnehézségi paraméterek között:  $r=0,55$ ,  $p<0,01$ .) Ezek az eredmények szintén azt a hipotézist igazolják, hogy bizonyos esetekben a tesztfelvevő elveszíti objektivitását, főként, ha az ingert kiemeljük a szokásos szöveggörnyezetéből.

A 3.16. ábra mutatja a tapasztalati összefüggés-megértés és a tapasztalati következtetés teszt itemnehézségi mutatóinak változását. A vonal alatt elhelyezkedő itemek könnyebbek voltak a hagyományos tesztfelvétel esetén, mindkét tesztben ezek az itemek voltak többségben.



Az itemnehézségi mutatók korrelációja magasnak bizonyult mindkét esetben (tapasztalati összefüggés-megértés  $r=0,75$ ,  $p<0,01$ ; tapasztalati következtetés  $r=0,76$ ,  $p<0,01$ .)



3.16. ábra

*A tapasztalati összefüggés-megértés és a tapasztalati következtetés test itemeinek viselkedésváltozása a hagyományos és számítógép-alapú adatfelvétel során*

#### 3.2.4. A papíralapú, kisiskolás diákok részére kidolgozott induktív gondolkodás teszt online változatával végzett mérések eredményei

A kisiskolás diákok részére kidolgozott 37 itemes induktív gondolkodás teszt belső konzisztenciája mind papír, mind számítógépes környezetben megfelelőnek bizonyult (Cronbach- $\alpha_{PP}=0,855$ ; Cronbach- $\alpha_{CB}=0,856$ ). A reliabilitásmutató értéke nem változott a közvetítő eszköz megváltozása hatására. Ennek oka lehet, hogy jelentős viselkedésbeni módosítás nélkül történt a papíralapú teszt digitalizációja.

Az induktív gondolkodás teszt mérési invarianciájának tesztelése során a metrikus invariancia modell modellilleszkedése azonos volt a konfigurális modell illeszkedési szintjével, sőt a legszigorúbb feltételek mellett sem történt változás az illeszkedés fokában (3.11. táblázat). A szigorú metrikus invariancia modell illeszkedése is azonosnak adódott a konfigurális invariancia modell illeszkedésével ( $\Delta CFI < 0,01$  és nem szignifikáns  $\chi^2$  teszt).

Ennek értelmében a közvetítő eszköz nem befolyásolta a diákok relatív helyzetét a mintában. Bármely két diák egymáshoz való viszonya független attól, hogy a két diák a tesztet papír vagy számítógép alapon oldotta-e meg (metrikus invariancia). Miután a teszt esetén a szigorú metrikus invariancia is teljesült, ezért nemcsak a diákok relatív helyzete, hanem a papír és a számítógépes környezetben elért nyerspontok és képességszintek is közvetlenül összehasonlíthatóak egymással, azaz a két környezetben történt képességszintbecslés azonos.

A diákok papíron és számítógépen elért átlagos teljesítménye se manifest, se látens szinten nem különbözött egymástól szignifikánsan [PP: átlag =47,52, sd=19,04; CB: átlag=45,99, sd=18,11; Cohen-d=0,08; látens változó CB: átlag: -0,10; se=0,06 (FF csoport

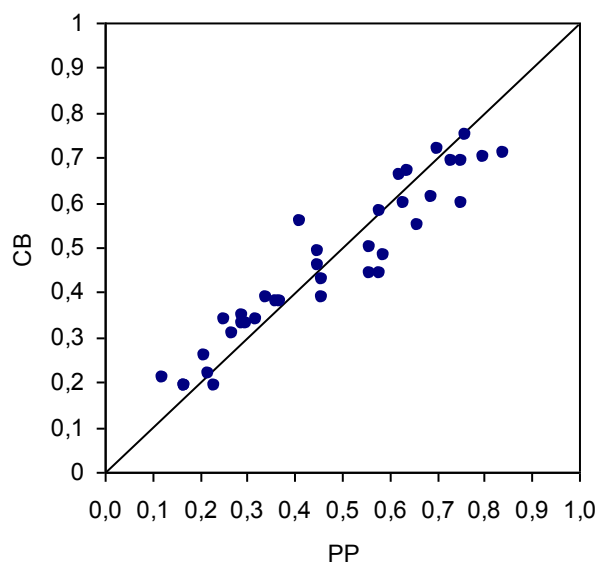
látens átlaga 0-ra állított);  $p=n.s.$ ; Cohen, 1988]. A teszt egészének viselkedését nem befolyásolta a kiközvetítés eszköze.

3.11. táblázat. Az induktív gondolkodás teszt mérési invarianciáját jellemző modellek illeszkedésmutatói

Modell	$\chi^2$	df	CFI	TLI	RMSEA	$\Delta\chi^2$ *	$\Delta df$ *	p
(1)	1791,2	908	0,929	0,923	0,037	-	-	-
(2)	1828,4	930	0,924	0,919	0,038	42,0	22	>0,01
(2.1)	1806,4	931	0,926	0,921	0,038	15,0	23	>0,05
(3)	1868,1	962	0,921	0,916	0,039	55,0	32	>0,05

Modell: (1) Konfigurális invariancia, (2) Metrikus invariancia, (2.1) Részleges metrikus invariancia, (3) Szigorú metrikus invariancia. \*  $\Delta\chi^2$  és  $\Delta df$  értékeket az MPlus DIFFTEST eljárásával számoltuk.

A teszt itemszintű viselkedésében rejlő azonosságok és eltérések feltérképezése céljából összevetettük ugyanazon feladat papíralapú és számítógép-alapú környezetben számolt nehézségi indexét (3.17. ábra). Egy pont egy feladatot reprezentál, a vízszintes tengelyen a feladat elhelyezése annak papíralapú környezetben mutatott nehézségi indexe alapján történt, míg a függőleges tengelyen a számítógép-alapú környezetben mutatott nehézségi indexe alapján. Ha egy feladat esetén a két index számszerűleg megegyezik, akkor a feladatot reprezentáló forma a diagonális vonalon helyezkedik el.



3.17. ábra

Az induktív gondolkodás teszt itemeinek nehézségi indexe papír (PP), illetve számítógépes (CB) környezetben

Az itemek elhelyezkedése általánosan a diagonálisan elhelyezkedő átlóhoz közeli, a nehézségi indexek korrelációja igen magas:  $r=0,948$ ,  $p<0,01$ , ami alátámasztja a korábbi

eredményt, miszerint a teszt PP és CB környezetben nagyon hasonlóan viselkedik. Közel azonosan méri a diákok induktív gondolkodásának fejlettségi szintjét – nemcsak a teszt egésze szintjén, de az egyes itemek szintjén is. A 3.17. ábra alapján megállapítható, hogy a teszt könnyebb itemei általánosságban számítógép-alapon kicsit könnyebbnek, míg a nehezebb itemek nehezebbnek bizonyultak, mint hagyományos tesztfelvételi környezetben.

E kutatási eredmény megerősíti az előző részben ismertetett eredményeket, miszerint kisiskolás diákok körében is sikeresen alkalmazhatóak a számítógép-alapú tesztek, sőt a diákok egér- és számítógéphasználati képessége fejlettségi szintjére vonatkozó eredményeket figyelembe véve – igény esetén – kidolgozhatóak olyan tesztek, melyek viselkedése azonos PP és CB környezetben. A közel azonos formátumban megjelenő tesztek jószágmutatóját, azaz az eredmények általánosíthatóságát és a tesztek működését nem befolyásolja a kiközzvetítés során alkalmazott médium.

### 3.2.5. A technológiaalapú tesztelés alkalmazhatósága az iskolakezdés kulcsfontosságú előkészítéseinek mérésére: összegzés, az elemzések korlátai, továbblépési lehetőségek

A kutatási eredmények összességében alátámasztották, hogy a számítógép-alapú tesztek már az iskolába lépéskor alkalmazhatóak több, az iskolai tanuláshoz nélkülözhetetlen előkészítések fejlettségi szintjének diagnosztizálása tekintetében a rendszeres oktatási gyakorlatban. Ezek a mérések gyakorlatilag bármikor, alacsony költséggel és minimális időráfordítással elvégezhetők.

Az elemzések korlátja, hogy az eredeti DIFER tesztek az óvodából az iskolába átlépő gyerekek mérésére tervezték, ezért a már iskolás gyerekeknél gyakran plafoneffektusba ütköztünk az elemzések során, ami nem kedvezett a tesztek működésének vizsgálatához. Ez kiküszöbölhető egy jövőbeni kutatással, melynek célpopulációját az óvodások alkotják, annak ellenére, hogy a számítógép-használat ebben a korban még további megvalósíthatósági tanulmányokat igényel (*Csapó, Molnár és Nagy, 2015*).

Az elemzések további korlátja, hogy az adatgyűjtést az eDia platform korábbi verziójával végeztük el, ami még csak a leggyakoribb itemformátumokat kezelte, de kevesebb lehetőséget adott új itemtípusok létrehozására. A korábbi nagymintás felmérésekhez képest a bemutatott kutatáshoz kisebb minta állt rendelkezésünkre, mivel még fejlesztés alatt levő rendszert használtunk. Az elemzési lehetőségeket szintén korlátozza a háttér-adatok hiánya. A kutatás jelenlegi fázisában nincs adatunk annak a lehetőségéről, hogy azonos eszközzel végzett ismételt teszteléssel kísérjük figyelemmel a fejlődést. (A DIFER korábbi személyes tesztelési módszerét rendszeresen használják erre a célra minden probléma nélkül.) Nincs adat továbbá a számítógépes tesztelés előrejelzési validitásáról sem.

Kimaradt a digitalizálásból az elemi számolási készség teszt sok lényeges feladata, mivel a tanulók szóbeli válaszait nem tudtuk rögzíteni. Ennek a veszteségnek az ára a gyengébb validitás és alacsonyabb megbízhatóság volt. A digitalizált változat tehát nem tekinthető az eredeti (magas megbízhatóságú és jó előrejelző érvényességű) teszt egyenértékű helyettesítésének, ezért további munka szükséges egy megfelelő számítógépes elemi számolási képességeteszt kidolgozása érdekében.

Szintén elmaradt az írásmozgás-koordináció teszt digitalizálása, ugyanakkor egy ahhoz közeli konstruktum (a finom kézmozgás) könnyen mérhető más technológiák felhasználásával, és egy alternatív konstruktum (billentyűzet és egérkezelés) szintén jól mérhető. Az idevágó

technológia rendelkezésünkre áll kutatási célokra, de az kérdéses, hogy az iskolák számára széles körben elérhető-e. A teszt viszonylag alacsony előrejelző validitása szintén felveti a kérdést, vajon megéri-e az energia-befektetést, amikor az új technológiákhoz szükséges hasonló pszichomotoros készségek mérése ígéretesebbnek tűnik.

Nem lehetett digitalizálni egy magas megbízhatóságú és jó előrejelző képességű, egyéni adatfelvételt igénylő tesztet, a szociális készségek (Zsolnai és Józsa, 2002) tesztet. A szociális készségek számítógépes mérésének elvi korlátja, hogy a gépi környezet mindig csak korlátozottan helyettesítheti a személyes interakciót. Mindemellett a szociális készség azon dimenzióira, amely az iskolakészültség szempontjából meghatározó jelentőségű: utasítások követésére való hajlandóság, feladatokra való folyamatos odafigyelés és a feladattartás, már lehet online teszteket készíteni.

A DIFER tesztek online változatával végzett kutatások burkolt célja egy olyan új tesztcsomag létrehozásának lehetősége feltérképezése volt, amely a legnagyobb mértékben kihasználja a számítógépes technológia nyújtotta előnyöket. Ennek az eszköznek a létrehozása és felhasználása még további kutatásokat igényel: (1) a jelenleg hagyományos módszerekkel mért területeket a számítógépes technológia által mérhetővé kell tenni, és olyan új konstruktumokat is kell mérni, amelyek különösen jól illenek a számítógépes technológiához; (2) tovább kell fejleszteni az online felmérési technológiát olyan funkciókkal, amelyeket az info-technológia más területein már használnak; (3) meg kell találni annak a módját, hogy a rendszeres kisgyermekkorú felmérések beépíthetők legyenek az oktatási folyamatba.

Számos technológiai megoldás létezik más területeken, amelyek beépíthetők az online mérési rendszerekbe azok lehetőségeit bővítve. Már most használható az interakció, szimuláció, a tárgyak irányítása a képernyőn, az ingerek új típusai, mint például a film és animáció. Az ingerek időzítése és az információ kontrollált megjelenítésének más módszerei szintén lehetségesek. Szintén egyszerűen megoldható a reakcióidő mérése, a billentyűzet gombnyomásának vagy az egér mozgásának rögzítése, bár további vizsgálatok szükségesek ahhoz, hogyan tudnak ezek az eszközök hozzájárulni a korai mérések gyakorlati problémáinak megoldásához.

További kutatások szükségesek ahhoz is, hogy megismerjük az online mérések alkalmazhatóságát az oktatásban. Az ökológiai érvényesség egyike azon kérdéseknek, amelyek alapos mérlegelést igényelnek. Az előrejelzési validitás kritikus fontosságú azoknál a teszteknél, amelyek a későbbi tanulás feltételeit mérik, és a későbbi problémák korai indikátorainak felismerésére használhatók. Az itt tárgyalt online tesztek vizsgálata a jelen tanulmány továbblépéseként már meg is kezdődött. Az ismételt tesztelés hatásainak feltárása szintén megkezdődött, de az elemzésekhez szükséges adatmennyiség összegyűjtése mindkét esetben hosszabb időt igényel. Mivel a tanulók felméréseinek adatait az eDia online felmérési rendszerben tároljuk, nem csak a tesztelési adatok, hanem a kérdőívekkel gyűjtött információk is egy helyre gyűjthetők, és a növekvő információs adatbázis lehetővé teszi azoknak a kérdéseknek a későbbi vizsgálatát, amelyekre most nem találtunk válaszokat.

A DIFER kidolgozása során meghatározó szempont volt, hogy azt az óvoda-iskola átmenet során többször fel lehessen venni, azzal a tanuló fejlődését nyomon lehessen követni. Az online diagnosztikus mérési rendszer is iskolakezdéstől átveszi a fejlődés követésének feladatait, ezért egy új iskolakészültség tesztrendszernek az iskolát megelőző évekre, és az iskolakezdés időpontjában alkalmazandó, az induló szintet feltérképező, az alapvető szűrési feladatokat ellátó készség-mérésre kell koncentrálnia.

Az itt bemutatott vizsgálatban arra törekedtünk, hogy egy hosszú fejlesztési múlttal rendelkező, a gyakorlatban is jól bevált mérőeszközt a lehető legpontosabban átültessünk az új technológiai alapra. Ez a törekvés ugyanakkor azzal is járt, hogy nem használtuk ki a technológia által kínált összes lehetőséget. Ezek az elemzések nem csupán bizonyították a technológia használhatóságát egy ilyen különleges mérési helyzetben, hanem megmutatták annak sokféle előnyét is, amit egy új generációs eszközrendszer kifejlesztése során figyelembe lehet venni (*Csapó, Molnár és Nagy, 2015*).

## 4. A SZÁMÍTÓGÉP-ALAPÚ TESZTELÉS ALKALMAZÁSA 9-19 ÉVES KORBAN: A PROBLÉMAMEGOLDÓ KÉPESSÉG MÉRÉSI LEHETŐSÉGEINEK VÁLTOZÁSA

### 4.1. A problémamegoldó képesség mérési lehetőségei

A problémamegoldó képesség az emberiség egyik legfontosabb, túlélését és fejlődését leginkább meghatározó képessége. Megvalósítja, hogy tetteink előtt gondolkodjunk, ami jelentős mértékben megnöveli az ember és környezete közötti interakció hatékonyságát, és biztosítja, hogy lehetőségeink szerint a legjobb döntést hozzuk meg az adott szituációban. Problémamegoldó képességünk lehetővé teszi, hogy olyan célokat, megoldási lehetőségeket is megvalósítsunk, melyeket különben nem érünk el.

A problémamegoldó képesség és mechanizmusainak sokféleségét jellemzi, hogy számos, különböző típusú problémával találkozunk életünk során (Molnár, 2013a). Vannak, amelyek ismerősek, amelyekhez hasonlóakkal korábban már találkoztunk, esetleg sikeresen megoldottunk, és vannak teljesen újak (Reeff, Zabal és Blech, 2006). Találkozunk könnyebben megoldható, és bonyolultabb, összetettebb megoldási módot kívánó problémákkal (Frensch és Funke, 1995). Léteznek kevésbé jól meghatározott, úgynevezett rosszul definiált, és vannak jól definiált, pontos cél(oka)t tartalmazó problémák (Schraw, Dunkle és Bendixen, 1995; Jonassen, 1997). A problémák egy részében nem változnak a rendelkezésre álló információk, míg dinamikus problémák esetén folyamatosan változó környezetben kell megtalálni az optimális megoldást. Összességében komplex, ugyanakkor az egész életünket jelentős mértékben meghatározó, sőt a 21. század gyorsan változó társadalmában egyre fontosabb szerepet betöltő képességről van szó. Mint a tanulás, a tanultak mindennapi életben való alkalmazásának egyik alapvető fontosságú képessége, az ezredforduló óta az egyik legtöbbet vizsgált gondolkodási képességévé vált (Molnár, Greiff és Csapó, 2013). Többek között kiemelt helyet foglal el a legnagyobb nemzetközi pedagógiai mérés-értékeléssel foglalkozó kutatásokban [pl.: OECD PISA mérések (OECD, 2004, 2010b), NAEP kutatások (Bennett, Persky, Weiss és Jenkins, 2007), ATCS21 projekt (Griffin, McGaw és Care, 2012)], valamint a 21. század kulcsfontosságú képességei közé sorolták (Scottish Qualifications Authority, 2003; European Parliament, 2006; Binkley, Erstad, Herman, Raizen, Ripley, Miller-Ricci és Rumble, 2012). Mérésére, fejlesztésére számos ország nagy hangsúlyt fektet, oktatási programjuk szerves részét alkotja (OECD, 2010a).

A disszertáció e fejezetének célja a mérés-értékelés módszertanában történő fejlődési tendenciák szemléltetése a problémamegoldó képességgel kapcsolatos empirikus kutatások és a megválaszolható kutatási kérdések körének bővülésén, változásán keresztül; külön hangsúlyt fektetve (1) a pedagógiai kontextusban történő mérési lehetőségekre; (2) a jelen kutatások előzményeként az elmúlt 10 év vonatkozó papíralapú tesztekre alapozó kutatásaink szintetizálására; (3) a számítógép-alapú tesztelés nyújtotta lehetőségek és igények ismertetésére. A fejezet második felében (4) több, egymásra épülő, különböző korosztályban megvalósuló harmadik generációs számítógép-alapú, dinamikus problémamegoldó képesség fejlődését összességében tág életkori intervallumban vizsgáló kutatás-sorozat céljairól, módszereiről és az alkalmazott mérőeszközök, problémák működéséről számolunk be.

#### 4.1.1. A problémamegoldó képesség mérésének elméleti háttere és mérési tendenciái

„Egész életünk problémamegoldás” (Popper, 1999), de hogyan tudnánk meghatározni a problémát, és modellezni a problémamegoldó képesség használatának folyamatát, a problémamegoldást kutatási perspektívából? Hogyan jellemezhető a problémamegoldó képesség, és mikor van szükségünk alkalmazására? A problémamegoldó képesség összetettségét mutatja, hogy kutatása közel 100 éves múlttal rendelkezik, egészen a 20. század elejéig, a Gestalt-pszichológiáig nyúlik vissza.

Karl Duncker, az alakléktan egyik fő képviselője 1935-ben publikálta először németül, majd 1945-ben angolul *On problem-solving* című könyvét (Duncker, 1945). Ugyanebben az évben jelent meg a Gestalt-pszichológia másik jelentős képviselőjének, Max Wertheimer *Productive thinking* című könyve (Wertheimer, 1945), illetve Pólya György *How to solve it* problémamegoldó képességgel foglalkozó könyvének első kiadása is (Pólya, 1945/1957). 1945 jelentős fordulóponthoz vezetett, fellendítette a problémamegoldással kapcsolatos kutatásokat. Az alakléktan képviselői az észlelés szerveződési folyamatainak elméletét kiterjesztették a problémamegoldó gondolkodás folyamataira, hangsúlyozták a problémáról és a probléma megoldásának menetéről kialakított reprezentáció közötti különbséget (Novik és Bassok, 2005), valamint felhívták a figyelmet arra, hogy mindkét reprezentáció a probléma nehézsége és a problémamegoldó előzetes tudásának fényében változik. Elméletük értelmében az emberek képesek átlátni egy probléma szerkezetét, majd a megoldás érdekében újrastrukturálni azt (Csapó és Molnár, 2012). Kísérleteikben tudásszegény problémaszituációkat alkalmaztak és a megoldási folyamat során létrejövő felismerésekre fókuszáltak (Frensch és Funke, 1995). Az 1950-es évek után, a Gestalt-pszichológia háttérbe szorulásával párhuzamosan a problémamegoldó kutatások is visszaszorultak, egészen 1972-ig.

Erre az időszakra tehető Pólya (1969) matematikai természetű problémamegoldásra vonatkozó munkássága, aki szakított a korábbi, a problémamegoldást lineáris folyamatként modellező megközelítéssel és egy ciklikus, illetve kétirányú folyamatként értelmezte azt. A problémamegoldó gondolkodás hazai vizsgálataiban sokáig kizárólag a Pólya-féle értelmezés dominált (Csapó és Molnár, 2012).

1972-ben Newell és Simon *Human problem solving* című könyve ismét előtérbe helyezte és fellendítette a vonatkozó kutatásokat. A Gestalt-pszichológusokkal ellentétben az emberi megismerést információfeldolgozásként értelmező irányzat képviselői a lépésenkénti problémamegoldó folyamat kutatása mellett érveltek. Fő céljuk olyan általános problémamegoldó stratégiák kidolgozása volt, amelyek számos, különböző probléma megoldása során alkalmazhatók. Munkájukra nagy hatással volt a kognitív pszichológia emberi információfeldolgozásra vonatkozó megközelítése és a számítógép-tudomány mesterséges intelligencia kutatásai. Vizsgálataik során tudásszegény, előzetes szakterületi tudást nem igénylő problémákra fókuszáltak, munkájuk eredményeként megalkották az Általános Problémamegoldót (GPS – *General Problem Solver*; Newell és Simon, 1972), az emberi problémamegoldást modellező számítógépes programot, ami jelentős mértékben fellendítette a problémamegoldó képességgel foglalkozó kutatásokat.

Az 1970-es évek közepén-végén Amerikában előtérbe kerültek az előzetes tudás szerepével foglalkozó kutatások, amelyek rávilágítottak a GPS gyengeségére (Novik és Bassok, 2005). Az Általános Problémamegoldó a vonatkozó területspecifikus előzetes tudást teljes mértékben figyelmen kívül hagyva modellezte az emberi problémamegoldó folyamatokat,

holott azt minden egyes probléma esetében jelentős mértékben befolyásolja az adott személy területre vonatkozó tudása. Ez a felismerés az 1970-es évek végén, az 1980-as évek elején számos területre (matematika, fizika, politikatudomány, sakkozás stb.) vonatkozó kutatást indukált, melyek egyöntetűen elvetették az általános problémamegoldó elmélet, modell kidolgozásának lehetőségét és a terület szakértőjévé válásának folyamatára koncentrálnak rávilágítottak az előzetes tudás szerepére. Utóbbira *Duncker* a matematikával kapcsolatban már 1945-ben felhívta a figyelmet.

Az amerikai kutatásokkal ellentétben, Európában már az 1970-es években is a valós élet problémáihoz hasonló felépítésű, számítógépes laboratóriumi problémákkal dolgoztak a kutatók. Az amerikai területspecifikus problémákkal szemben a valós élet problémáihoz hasonlóan relatív komplex, szemantikailag gazdag (*Frensch és Funke, 1995*) és a kísérleti személyeknek új, több területet átfogó (területfüggetlen) problémák alkalmasnak bizonyultak a problémamegoldás folyamatának általános jelenségként való vizsgálatára.

Az ezredfordulóig a hazai problémamegoldással kapcsolatos empirikus kutatások alapvetően pilot jellegű vizsgálatok voltak. A problémák jellemzően különböző tudományterületekhez kötődtek. Matematikai természetű problémák megoldására irányultak *Kontra* (1996) kutatásai, biológiához kötődő problémák megoldási sikerességét vizsgálta *Revákné Markóczy* (2001); míg mind matematikával, mind természettudományokkal kapcsolatos problémákkal foglalkozott *Molnár* (2001) kutatásaiban.

Az ezredforduló gyors társadalmi-gazdasági változásai, a megváltozott munkaerő-piaci igények előtérbe helyezték a problémamegoldó képességgel kapcsolatos kutatásokat (*Csapó és Molnár, 2012*), illetve a problémamegoldó képesség iskolai kontextusban való felmérését. Fontosságát mutatja, hogy az OECD PISA vizsgálat sorozatában helyet kapott a negyedik ciklusban, diszciplínákat átmetsző, a tudás alkalmazhatóságát vizsgáló területként, s szerepelt a második (komplex problémamegoldás – 2003; *OECD, 2004*), az ötödik (kreatív problémamegoldás – 2012; *OECD, 2010b; Funke, 2010*) és a hatodik ciklusban (kollaboratív problémamegoldás – 2015) is. Mindemellett a nemzeti és a nemzetközi 21. századi készségek és képességek definiálását célzó programok közel egyöntetűen a legfontosabb 21. századi képességek közé sorolták a problémamegoldó képességet (*Binkley és mtsai, 2012*). Ennek ellenére nincs egységes, mindenki által elfogadott meghatározás a területen. Sőt a PISA mérések keretein belül 2003, 2012 és 2015-ben végzett és tervezett kutatások problémamegoldó modelljei is jelentős mértékben különböznek egymástól, más-más fókuszról vizsgálják a 15 éves diákok problémamegoldó képességének fejlettségi szintjét.

Miután a problémamegoldó képesség fejlettségi szintjének nagymintás mérésére nemzetközi szinten először a PISA kutatások keretein belül vállalkoztak a kutatók. Ezért a továbbiakban a problémamegoldó képesség meghatározásának változását a PISA kutatások elméleti alapját adó elméleti keretrendszerre építjük. Elfogadtuk és kiindulási pontként vettük azon kutatási eredményeket (*Sternberg, 1994; Novik, Hurley és Francis, 1999; Funke, 2001, 2010*), miszerint függetlenül a megoldandó probléma természetétől (pl.: komplex, interaktív vagy kollaboratív) a problémamegoldó képességet számos általános gondolkodási képességünk is meghatározza. Releváns kutatási irány, ha általánosságban, és nem csak problémátípus-specifikusan beszélünk a problémamegoldó képességről, mely kutatások rövid történeti kitekintését l. *Molnár* (2013a) tanulmányában.

A 2000-ben induló, ma már több mint 70 ország több mint félmillió 15 éves diákjának részvételével zajló OECD PISA kutatássorozat korábban nem tapasztalt fejlődést indukált a



pedagógiai mérés-értékelés területén, így a problémamegoldó képesség kutatásában is. A PISA-felmérés nem vállalkozott és nem is vállalkozhatott a korábbi kutatások, elméleti modellek szintetizálására, a problémamegoldó képesség korábban vizsgált dimenzióinak egyesítésére. A mérések alapját képező elméleti keretrendszer célja egy világszerte elfogadható, papíralapú méréssel kultúrafüggetlenül megvalósítható, ugyanakkor a kognitív tudományok legújabb kutatási eredményeire építő és a 21. század igényeihez illeszkedő modell kidolgozása volt. A 2003-ban újonnan kidolgozott elméleti keret fő vázát ezért *Pólya György* mai napig széles körben ismert és elfogadott problémamegoldás-modellje alkotta (*Pólya*, 1957, 1969). A modell egyes részei, a problémamegoldás folyamatának fő fázisai a későbbi (2012 és 2015) kutatások elméleti keretrendszerében is megtalálhatóak: a probléma azonosítása, megértése, reprezentációja, mint a tudás elsajátítása, a probléma megoldása és az eredmények kommunikálása, mint a tudás alkalmazása.

A 2003-as PISA kutatás alapját adó meghatározás értelmében a problémamegoldás az „egyén képessége arra, hogy kognitív eljárásokat használjon olyan reális, diszciplínákat átmetsző helyzetekben, amikor a megoldáshoz vezető út nem válik azonnal nyilvánvalóvá, [...] tartalmak, amelyek esetleg felhasználhatóak, nem találhatók meg a matematika, az olvasás vagy a természet-tudomány egyetlen területén belül” (*OECD*, 2003, 156.). A kutatás során – a papíralapú mérés adta korlátok között – alkalmazott problémák kontextusa eltávolodott az iskolában megszokott szöveggörnyezettől, és helyet kaptak a mindennapi élettel, a munkával és a szórakozással kapcsolatosan felmerülő problémák. A problémák megoldása során felhasználandó tudásterületek kielégítve az életszerűség és a valóság kritériumait is nem kizárólagosan egy-egy speciális tudományterületről, hanem tudományterületeket összekötve, számos forrásból származtak, tartalmaztak elemeket (*Csapó*, 2005; *Molnár*, 2006b). A kutatás utólagos korlátjaként fogalmazódott meg, hogy számos diák interaktívan is kipróbálta volna a problémahelyzetek adta különböző lehetőségeket, alkalmazva a problémamegoldó képesség működését jellemző próba és hiba (*trial and error*) stratégiát, azonban erre papíralapon nem volt lehetőség (4.1. ábra). A papíralapú, statikus problémák segítségével történt kutatások eredményei egyre kevésbé tükrözték azt, hogy a mai kor diákja milyen jó problémamegoldónak bizonyul a technológiai eszközökkel telített mindennapi életében. Ennek hatására a problémamegoldó képesség újabb dimenzióinak kutatását lehetővé téve a 2012-es adatfelvétel már számítógép-alapon valósult meg.

Annak ellenére, hogy ez a korábban nem tapasztalt mennyiségű interaktivitás és dinamikuság mesterséges környezetből fakad, a körülöttünk lévő, hihetetlen gyorsasággal fejlődő technológiai eszközök és programok okozzák, e környezet mérés-elméleti szempontból történő modellezése jelentős feladatot és kihívást ró és rótt a kutatókra. *Buchner* (1995) az interaktív problémamegoldó képesség mérésének két elfogadott megközelítését foglalja össze: számítógépen szimulált, a valós élethez hasonlóan számos változót tartalmazó mikrovilágok [l. pl.: *Dörner* 1980 híres Lohhausen problémája 2000-nél több változót tartalmaz, l. *Dörner*, *Kreuzig*, *Reither* és *Stäudel*, 1983; *Gardner* és *Berry*, 1995 orvosi rendszert modellező problémája; *Moray*, *Lootsteen* és *Pajak*, 1986 ipari termeléses mikrovilága], vagy mesterséges, egyszerűsített, bizonyos szabályokat követő, de mégis összetett probléma, minimális komplex rendszer (*minimal complex system*) alkalmazása [pl.: a lineáris egyenletekre épülő DYNAMIS megközelítés, l. *Funke* (1992)].

Pszichometriai szempontból az adatfelvételhez szükséges idő, a probléma valódisága, életszerűsége, a mérés validitása, reliabilitása, a problémák skálázhatósága, a mögöttes

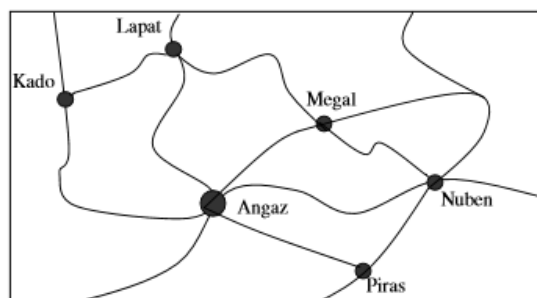
méréselméleti modell tekintetében mindkét lehetőség számos előnnyel és hátránnyal bír (részletesen l. *Greiff, 2012*). Míg az előbbi jobban leképezi a mindennapi élet problémáit, addig az utóbbi méréselméleti szempontból kedvezőbb tulajdonságokkal írható le. Például a Lohhausen probléma nehézségi szintje nehezen skálázható, adatfelvétele sok órát vesz igénybe, ellenben a DYNAMIS-megközelítésen alapuló mesterséges probléma néhány perc alatt megoldható, nehézség tekintetében könnyen skálázható. Előbbiből nehézkes akár egy teszt összeállítása is, miután egy probléma megoldása is sok órát vesz igénybe, a probléma egyes részeinek megoldása pedig függ egymástól, utóbbi esetén számos problémát tartalmazhat például egy iskolai tanóra alatt megoldható teszt. A két megközelítés közös tulajdonsága, hogy számítógép-alapú teszteléshez kötött.

## VAKÁCIÓ

Ebben a feladatban egy nyaralás útvonalát kell megtervezned.

Az 1. és 2. ábrán az adott terület térképe és a városok közötti távolságok láthatók.

1. ábra: Térkép a városokat összekötő útvonalakról



2. ábra: Az egyes városok közötti legrövidebb útvonal kilométerben

Angaz						
Kado	550					
Lapat	500	300				
Megal	300	850	550			
Nuben	500		1000	450		
Piras	300	850	800	600	250	
	Angaz	Kado	Lapat	Megal	Nuben	Piras

### 1. feladat: VAKÁCIÓ

X602Q01 – 0 1 9

Számítsd ki, hogy milyen hosszú a legrövidebb útvonal Nuben és Kado között!

Útvonal hossza: ..... kilométer.

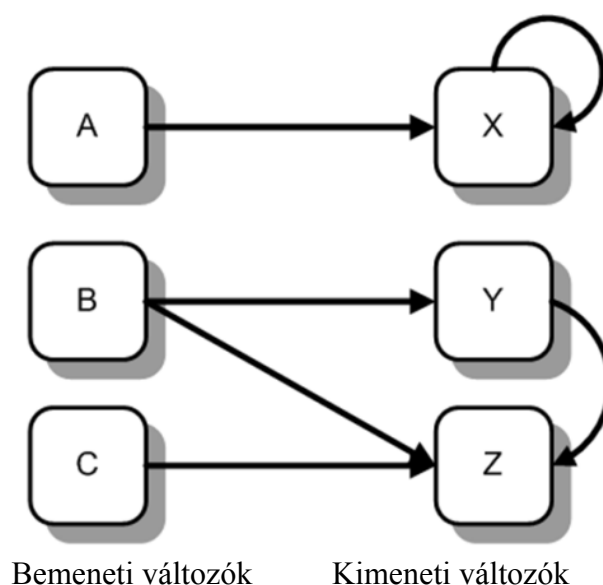
#### 4.1. ábra

*Példafeladat a PISA 2003-as Problémamegoldás moduljából (papíralapú)*  
(Forrás: [oktatas.hu](http://oktatas.hu), az Oktatási Hivatal honlapja)

A PISA-kutatásokban a DYNAMIS-megközelítésre alapozó MicroDYN-modell került adaptálásra, amely alapvetően nagymintás, számítógép-alapú tesztelést alkalmazó pedagógiai kutatások kivitelezésére került kidolgozásra. Az elméleti keretrendszernek megfelelő interaktív, dinamikusan változó, korlátozott mennyiségű változót (maximum három nem valós, kitalált nevekkal ellátott bemeneti – A, B, és C – és három kimeneti – X, Y és Z l. 4.2. ábra) tartalmazó, jól meghatározott, a tesztelt személy számára előre ismeretlen relációkkal,

függvényekkel leírható fiktív kontextusú problémahelyzeteket tartalmaz, amelyek 3-4 perc alatt megismerhetők, azonosíthatók, felfedezhetők (tudáselsajátítás fázisa), majd utána a megadott cél elérése érdekében működtethetők (tudásalkalmazás fázisa). Kontextusát tekintve minden egyes probléma más történet köré épül.

Az adatfelvétel során a problémamegoldók csak a bemeneti változók értékét manipulálhatták (Wüstenberg, Greiff és Funke, 2012; Greiff és Funke, 2010), aminek hatására a probléma háttérében lévő összefüggésrendszer szerint változtak a kimeneti értékek. Megfelelő stratégia használatával felfedezhető a változók közötti kapcsolatrendszert, amit a probléma alatt található, a bemeneti és a kimeneti változókat megjelenítő modellen meg is kell, hogy jelenítsen a tesztet megoldó személy. A probléma megoldásának második fázisában a problémamegoldónak működtetnie is kellett ugyanazt a rendszert. A kimeneti változók előre meghatározott célértékét kell elérnie négy lépésben – kutatástól függően – 90 vagy 180 másodpercen belül. A problémamegoldás e fázisában (tudás alkalmazása) a program megjeleníti a problémamegoldó számára a helyes összefüggésrendszert.



4.2. ábra

*Egy tipikus MicroDYN probléma szerkezete három bemeneti (A, B, C) és három kimeneti (X, Y, Z) változóval, valamint különböző típusú hatásmechanizmusokkal (egy és többszörös hatás, egyszeres és többszörös függés, sajátdinamika és mellékhatás)*  
 (Forrás: Greiff, Wüstenberg, Molnár, Fischer, Funke és Csapó, 2013)

Ezzel a modellel a 2012-ben megvalósuló PISA kutatásban a számítógép-alapú tesztelés adta lehetőségek kihasználásával kiküszöbölték a 2003-as kutatás egyik fő korlátját, az interaktivitás hiányát. A szakértői csoportok által kidolgozott definíció értelmében az interaktív, kreatív problémamegoldó képesség az a képességünk, ami lehetővé teszi, hogy ismeretlen technológiai eszközökkel interakcióba lépve, képesek legyünk azok működését – bizonyos célok elérése érdekében – feltérképezni (OECD, 2010c). A kutatás során alkalmazott problémák már egyetlen tudományterülethez sem kötődtek, területáltalános, és a tesztelt személy interakciója hatására statikus helyzetekből álló dinamikus változó problémák voltak

(4.3. ábra). A számítógép-alapú tesztelés adta lehetőségek további kihasználási területe az adatfelvétel során keletkezett logfájlok elemzése, amely elemzésektől még részletesebb, a problémamegoldás folyamatának alaposabb elemzése és megértése várható (l. 5. fejezetet).

A 2015-ben esedékes PISA kutatás a munkaerőpiaci igényeket követve már nem egyéni, hanem társas környezetben vizsgálja a diákok problémamegoldó képességének fejlettségi szintjét. Megtartja a 2012-es elméleti koncepció interaktív- és technológiaorientáltságát (Greiff, Holt és Funke, 2013) kiegészítve eddig nem vizsgált nem-kognitív képességekkel, úgynevezett szociális képességekkel. Az ATCS21 projekt eredményeire (Griffin, McGaw és Care, 2012) is építő elméleti keretrendszer és a kutatás során kiközvetítésre kerülő problémák még kidolgozás alatt állnak, ugyanakkor az előzetes empirikus eredmények arra utalnak, hogy a problémamegoldó képesség ezen kollaboratív-szociális dimenzióból is mérhető, és fejlettségi szintje számszerűsíthető (Greiff és mtsai, 2013; OECD, 2013d; l. 4.4. ábra).

**CLIMATE CONTROL**

You have no instructions for your new air conditioner. You need to work out how to use it.

You can change the top, central and bottom controls on the left by using the sliders ( $\rightleftharpoons$ ). The initial setting for each control is indicated by  $\blacktriangle$ .

By clicking APPLY, you will see any changes in the temperature and humidity of the room in the temperature and humidity graphs. The box to the left of each graph shows the current level of temperature or humidity.

Top Control:  $-- - \blacktriangle + ++$

Central Control:  $-- - \blacktriangle + ++$

Bottom Control:  $-- - \blacktriangle + ++$

Temperature: 25

Humidity: 25

APPLY RESET

**Question 1: CLIMATE CONTROL CP025Q01**

Find whether each control influences temperature and humidity by changing the sliders. You can start again by clicking RESET.

Draw lines in the diagram on the right to show what each control influences. To draw a line, click on a control and then click on either Temperature or Humidity. You can remove any line by clicking on it.

Top Control

Central Control

Bottom Control

Temperature

Humidity

?  $\rightarrow$

4.3. ábra

Példafeladat a PISA 2012-es Problémamegoldás moduljából (számítógép-alapú)  
(Forrás: OECD, 2014. 37-38. o.)

Angliában a NAEP (*National Assessment of Educational Progress*) kutatások keretein belül került sor a problémamegoldás technológiailag gazdag környezetben történő mérésére (PS in TRE – *Problem Solving in Technology-Rich Environments*; Bennett, Persky, Weiss és Jenkins, 2007). Mindennek hatására jelentős mértékben átalakult a problémamegoldó képesség, mint konstruktum értelmezése, meghatározása is. A problémamegoldó képesség vizsgálatával kapcsolatos kutatásokban kevésbé merült fel az igény arra, hogy ugyanazon konstruktum mérése valósuljon meg technológialapon, mint korábban papíralapon, sokkal inkább a technológia adta lehetőségek minél szélesebb körű kihasználásával eddig nem vizsgált részképességeinek kutatása, illetve maga a konstruktum, mint a 21. század egyik kulcsfontosságú képessége, került újradefiniálásra.

**PISA 2015 Unit name: The Aquarium** Time remaining: 17 minutes ? ➡

**Task 1 of 7**  
You and Abby have 3 minutes to decide how you will find the best conditions for the fish to live in the aquarium. Start with chatting to Abby.

**CHAT**

You: I'll try to work with my control panel

Abby: Wait – let me share my control panel with you first. Can you see it? Click on "Share it" so I'll see yours

You: Cool! Now it'll be easier.

Abby: What should we do now?

You:
 

- Are you ready to start?
- Let's play with the control panel
- Let's change the scenery

**Control panel**

Water type: ☐ Fresh ☒ Sea  
 Scenery: ☐ Rocky ☒ Plants  
 Lighting: ☐ Low ☒ High

**Abby's control panel**

Food type: ☐ Dry ☒ Food blocks  
 Fish: ☐ Few ☒ Many  
 Temperature: ☐ Low ☒ High

Tryout conditions

**Results**

Bad OK Great

#### 4.4. ábra

*Példafeladat a PISA 2015-ös Problémamegoldás moduljából (számítógép-alapú)*  
 (Forrás: OECD, 2013d. 55. o.)

Összefoglalóan, hasonlóan a mindennapi élet problémáinak változatosságához, a problémamegoldó képesség mérése pedagógiai kontextusban is számos perspektívából valósítható meg. Az alkalmazott problémák a probléma kontextusa szerint lehetnek területspecifikus vagy konkrét tudományterülethez nem kötődő problémák (Molnár, 2013a). A problémahelyzet természete szerint lehetnek ismerős vagy teljesen új problémák (Reeff, Zabal és Blech, 2006), lehetnek könnyebben megoldható vagy komplexebb, bonyolultabb megoldási

módot kívánó problémák (*Frensch és Funke, 1995*), lehetnek kevésbé jól definiált vagy több jól definiált célt tartalmazó problémák (*Schraw, Dunkle és Bendixen, 1995; Jonassen, 1997*), illetve lehetnek statikus, vagy időben dinamikusan változó, interaktív problémák (*OECD, 2010c*). Az adott problémán (együtt) dolgozók száma szerint elkülönítjük az egyéni, illetve a társas, csoportban történő, kollaboratív problémamegoldó képességet. Ahogy a kutatások alapját jelentő modellek, meghatározások, megközelítések és mérési lehetőségek változnak, úgy módosulnak az alkalmazott problémák és problémahelyzetek is. Fokozatosan háttérbe kerültek a hagyományos adatfelvételen alapuló, statikus problémákat tartalmazó kutatások és a technológia elérhetővé válásával a nemzetközi nagymintás kutatásokban is megtörtént az átállás a számítógép-alapú adatfelvételre, kihasználva a technológia adta lehetőségek széles körét.

#### 4.1.2. A területspecifikus problémamegoldó képesség mérése papíralapú tesztekkel: a hazai kutatások szintetizálása

Az ezredfordulóig a hazai kutatások alapvetően papíralapú tesztekre alapozó, egy-egy tudományterülethez kötődő problémákat alkalmazó, pilot jellegű vizsgálatok voltak. A nemzetközi nagymintás mérések hatására az ezredforduló után hazánkban is fellendültek a problémamegoldó képesség fejlődésével, az azt befolyásoló tényezők feltérképezésével és a fejlesztés lehetőségeivel kapcsolatos kutatások (l. pl. *Molnár, 2002, 2006b, 2010b, 2012, 2013a; Revákné Markóczi, Tóth és Tóthné Kosztin, 2009; Molnár, Greiff és Csapó, 2013*). A hazai kutatások, hasonlóan a nagy nemzetközi vizsgálatok adatfelvételeihez, az ezredforduló első tíz évében alapvetően papíralapú tesztekkel történtek, statikus és területspecifikus problémákat alkalmazva osztálytermi környezetben, az egyéni képességfejlettséget vizsgálva.

A továbbiakban először a szegedi műhelyhez köthető, területspecifikus komplex problémamegoldó képesség fejlődését középpontba állító, 2001 és 2011 között zajló kutatások elméleti keretrendszerét és felépítését tekintjük át dióhéjban. Az ezredforduló után induló kutatássorozat keretein belül a problémamegoldó képesség vizsgálatának egy új szemléletű modellje, és az arra építő, mára már 100 problémát tartalmazó tesztrendszere került kidolgozására.

A kutatás keretrendszere integrálta mind az amerikai, mind az európai, valamint a 2000-ben induló OECD PISA vizsgálatok műveltségkonceptiójának szemléletét. Az újonnan kidolgozott, korábban innovatívnak számító keretrendszernek megfelelően a kutatás problémái matematikai és/vagy természettudományi ismeretekkel megoldható, életszerű kontextusba ágyazott (pl. iskolai vagy családi kirándulás, pizzarendelés, vásárlás), szemantikailag gazdag problémák voltak, melyek tág életkori intervallumban tették lehetővé a diákok alkalmazható tudásának papíralapú tesztekkel történő vizsgálatát.

A kutatások felépítése a valószínűségi tesztelmélet adta eszközrendszer felhasználásával lehetővé teszi, hogy az elmúlt tíz év kutatási eredményeit egyesítve közös képességskálán jellemezzük a 3–11. évfolyamos diákok problémamegoldó képességének fejlődését, a különböző típusú problémák megoldási sikerességének alakulását, valamint nem, iskolatípus és szocioökonómiai faktor mentén összehasonlítsuk a képzett részminták teljesítményének változását. A disszertáció ezen alfejezetében ismertetett elemzések célja 3–11. évfolyamos diákok (1) területspecifikus problémamegoldó képességének, fejlettségi szintjének meghatározására alkalmas itembank kidolgozása, (2) területspecifikus problémamegoldó

gondolkodás képesséskálájának kialakítása, (3) a különböző típusú problémák megoldása során mutatott fejlődésbeli különbségek és azonosságok leírása, valamint (4) a nem és a szülő iskolai végzettségének teljesítménybefolyásoló hatásában történt változás azonosítása tíz év távlatában.

A fejezetben bemutatott elemzésekbe hat nagymintás empirikus vizsgálat adatbázisát, azaz közel 24000 diák adatát vontuk be (4.1. táblázat). Az első adatfelvétel 2002-ben zajlott 3–11. évfolyamos nagyvárosi diákokkal ( $n=4890$ , kb. 500/kohort). Ezt követte 2004-ben egy hátrányos helyzetű, 3–8. évfolyamos diákok körében végzett kutatás ( $n=6054$ , kb. 1000/évfolyam; a mintában a nehéz szociális helyzetben lévő tanulók és a roma diákok aránya magasabb az országos átlagnál, az adatfelvételben érintett iskolák közel felében 30 százalék feletti a halmozottan hátrányos helyzetű tanulók aránya); majd 2006-ban egyrészt a 2004-es adatfelvétel 5. évfolyamos diákjainak követéses vizsgálata valósult meg ( $n=937$ ), másrészt országos reprezentatív mintán 7. és 11. évfolyamos diákok problémamegoldó képességének feltérképezése ( $n=5827$ ). 2010-ben 7. évfolyamos, régió és településtípus szerinti reprezentatív mintán ( $n=3572$ ) végeztünk adatfelvételt, majd 2011-ben ismételtén 3–11. évfolyamos diákok problémamegoldó képességének fejlettségét vizsgáltuk. A 2011-es mintát a szülő iskolai végzettsége szerint illesztettük korábbi országos reprezentatív minta adataihoz, és az elemzésből kizártuk azon mérések és évfolyamok eredményeit, ahol az adott teszt reliabilitásmutatója (Cronbach- $\alpha$ ) 0,71 alatti volt. Ez a kritérium a 2011-es mérés két évfolyamát (9. és 10.) és az adott teszteken belül több itemet is érintett. Ennek következtében a 2011-es mérés tekintetében az illesztés és kizárás miatt összesen 2642 diákra vonatkozó adat került az egyesített adatbázisba.

Az anya iskolai végzettségét mindegyik mérés esetében hatfokú skálán jellemeztük (1: nem végezte el az általános iskolát, 2: általános iskola, 3: szakiskola, 4: érettségi, 5: főiskola, 6: egyetem), átlagos értéke az integrált mintán 3,6. A 2002 és 2011 közötti időszakban megfigyelhető az anya iskolai végzettsége átlagának növekedése. Míg 2002-ben az általános iskolás diákok anyáinak átlagos iskolai végzettsége 3,24, addig ez a szám (4,06) 2011-ben szignifikánsan magasabb (4.1. táblázat).

A közel tíz évvel ezelőtt induló kutatássorozat keretein belül a problémamegoldó képességgel kapcsolatban egy új modell és párhuzamosan egy új szemléletű tesztcsomag kidolgozására került sor. A modell integrálja az amerikai és az európai irányzatok szemléletét, valamint a PISA vizsgálatok műveltségkonceptióját. A modell értelmében kidolgozott tesztek a matematika és a természettudományok területén életszerű, szemantikailag gazdag környezetben vizsgálják a diákok alkalmazható tudását, valamint tág életkori intervallumban és iskolai kontextusban valószínűsítik meg a diákok képességvizsgálatát.

A valószínűségi tesztelmélet adta eszközöket kihasználva az egyes keresztmetszeti vizsgálatok tesztjeit úgy dolgoztuk ki, hogy azok – mind az adott mérés keretein belül szereplő különböző nehézségű tesztek, mind a különböző adatfelvételek során alkalmazott tesztek – horgonyitemek segítségével összekapcsolhatóak legyenek. A mérések során összesen 100 jól működő item kidolgozása és paraméterezése valósult meg, az elemzésekből kizártuk a kevésbé jól működő itemeket. Az alkalmazott itemek között kezdetben nagyobb arányban fordultak elő horgonyitemek, majd a mérések előrehaladtával egyre több új item kidolgozása valósult meg. A 2002-es adatfelvétel 54 iteme közül később 40 item került bele újabb tesztekbe, a második mérés 40 iteme közül 2 volt teljesen új item, a többi 38 már a 2002-ben alkalmazott itemek közül került ki. A 2006-os adatfelvételnél a horgony- és a nem horgonyitemek aránya 17:9,

2010-ben 17:11, végül 2011-ben 19:19, azaz az itemek fele már teljesen új, korábban még nem alkalmazott item volt. A horgonyitemek között vannak kövér horgonyok (*fat anchor*; 1. korábban, 2.2.1.1. rész) is, amelyek minden egyes adatfelvétel során szerepeltek a tesztek feladatai között, illetve a többi horgonyitem kiosztása során külön figyelmet fordítottunk arra, hogy bármely adatfelvétel bármely másik adatfelvétel során alkalmazott teszttel horgonyozható legyen.

4.1. táblázat. Az adatelemzésekbe bevont kutatások mintáinak főbb tulajdonságai

Adatfelvétel éve	Évfolyam	N	Főbb tulajdonság	Anyá átlagos iskolai végzettsége (adatok hány százalékában ismert)
2002	3.	543	nagyvárosi diákok	3,51 (72,0)
	4.	558		3,28 (82,0)
	5.	515		3,15 (87,6)
	6.	556		3,28 (91,5)
	7.	498		3,18 (87,6)
	8.	491		3,03 (91,0)
	9.	614		3,22 (93,2)
	10.	590		3,22 (94,1)
	11.	525		3,26 (92,6)
2004	3.	1015	hátrányos helyzetű diákok	n. a.
	4.	1029		n. a.
	5.	998		3,82 (79,9)
	6.	1047		3,56 (88,3)
	7.	1053		3,44 (92,0)
	8.	912		3,60 (90,2)
2006	7. (követéses)	937	hátrányos helyzetű diákok	3,82 (70,6)
	7.	3451	országos	3,83 (91,0)
	11.	2376	reprezentatív	3,94 (77,8)
2010	7.	3572	országos reprezentatív	3,70 (73,6)
2011	3.	309	illesztett országos reprezentatív	3,95 (61,8)
	4.	373		4,32 (47,2)
	5.	470		3,97 (58,7)
	6.	432		4,02 (61,1)
	7.	520		4,06 (78,5)
	8.	320		4,06 (82,5)
	11.	218		3,92 (100,0)
2002–2011 együtt	3-11.	23922	-	3,65 (74,8)



A problémamegoldó képesség tesztjein életszerű helyzetekbe (pl. iskolai vagy családi kirándulás, pizzarendelés, vásárlás) ágyasztuk az alapvetően matematikai eszközökkel megoldandó problémákat (l. 4.5. ábra). Formailag a feladatlap oldalait két részre osztottuk: bal oldalon találhatóak az információk realiztikus formában (pl. hirdetés, térkép, pizzaárak), jobb oldalon a problémák egy történetbe ágyazva. A felmerülő problémák alapvetően három csoportba sorolhatók: (1) a megoldásához szükséges információt tartalmazza a feladatlap, ám nem a megszokott iskolai, hanem életszerű formában; (2) a megoldásához nem minden információ adott, de a hiányzó információk a tananyag részét képezik; (3) a megoldásához nem található meg minden háttérinformáció, azonban azokkal a diákok, ha nem is az iskolai tanórán, de a hétköznapi életben találkozhattak. A feladatlapokon szereplő és az elemzésekbe bevont problémák közül 28 sorolható az első, 47 a második és 25 a harmadik kategóriába.

Apród pizzák				
Ár	20 cm	28 cm	32 cm	38 cm
Sajtos pizzaszósz, sajt, parmezánmarinka	rendel	rendel	rendel	rendel
Sonkás pizzaszósz, sonka, sajt	rendel	rendel	rendel	rendel
Gombás pizzaszósz, gomba, sajt	rendel	rendel	rendel	rendel
Kolbászos pizzaszósz, kolbász, sajt	rendel	rendel	rendel	rendel
Szalámi pizzaszósz, szalámi, sajt	rendel	rendel	rendel	rendel
Caribi pizzaszósz, sonka, trópusi gyümölcsök, sajt	rendel	rendel	rendel	rendel
Erdő kapitánya pizzaszósz, sonka, gomba, sajt	rendel	rendel	rendel	rendel
So-lu pizzaszósz, sonka, kukorica, sajt	rendel	rendel	rendel	rendel

Főnömes pizzák				
Ár	20 cm	28 cm	32 cm	38 cm
Hawaii pizzaszósz, sonka, ananász, kukorica, sajt	rendel	rendel	rendel	rendel
Son-go-lu pizzaszósz, sonka, gomba, kukorica, sajt	rendel	rendel	rendel	rendel
Spencer tejföl, fűt tojás, bacon, mozzarella	rendel	rendel	rendel	rendel
Mexikói chili szósz, sonka, bab, kukorica, pufferoni paprika, sajt	rendel	rendel	rendel	rendel

**KOSÁR TARTALMA**

Üres a kosár

Összesen: 0 Ft

\* csomagolási és száll. díj

Név: \_\_\_\_\_

Város: \_\_\_\_\_

Cím: \_\_\_\_\_

Telefon: \_\_\_\_\_

E-mail: \_\_\_\_\_

Megjegyzés: \_\_\_\_\_

☒ jegyzd meg a rendelésem

**AKCIÓ!!!**

**EGYET FIZET, KETTŐT KAP.\*\***

\* Kiszállítási díj: 180 Ft. A doboz ára 70 Ft.

\*\* Az ajánlat kizárólag a 28cm-es pizzára vonatkozik.

Másnap délelőtt ájtott négy haverom, 11-kor már nagyon éhesek voltunk, rendeltünk egy-egy pizzát. Anna és Juli közösen kértek egy sonkás pizzát, a fiúk pedig egy-egy kicsi gombásat, én egy közepes mexikóit. Ez vajon mennyibe kerülhetett?

A: 3360Ft      B: 3570Ft  
C: 3780Ft      D: 2460Ft

A pizzafutár hozott egy 20%-os engedményre jogosító ajándékkupont is. Miután megérkezett a pizzánk, olyan jó illata volt, hogy anyuék is úgy döntöttek, ők is pizzát esznek ebédre. Apa betült a kocsiba, vitte a kupont és hozott magának egy nagyobb Erdő kapitánya pizzát, anyának pedig egy kicsit kisebb ananászos pizzát. Mennyi pénzt kellett apunak legalább vinni a pizzériába?

A: 1400Ft      B: 1700Ft  
C: 2100Ft      D: 2400Ft

Egész jól jártak a végén!

#### 4.5. ábra

#### Példafeladat a problémamegoldó feladatlapról

Az eredmények közös képességskálára konvertálásánál, az adatok skálázásánál dichotóm adatok elemzésére alkalmas egyparaméteres modellt, a Rasch modellt alkalmaztuk. Az egyéni képességszintek meghatározása során WLE (*weighted likelihood estimation*), míg a populációparaméterek meghatározása során PV (*plausible values*) értékeket számoltunk (ezek leírását és értelmezését l. Molnár, 2013b könyvében), majd a logitskálán lévő képességszint-értékeket a 8. évfolyamos diákok átlagos teljesítménye alapján lineáris transzformációval 500 pont átlagú és 100 pont szórású skálára transzformáltuk. A problémátípusonként meghatározott képességszintek transzformációja esetén ugyanezekkel a paraméterekkel dolgoztunk.

A kidolgozott 100 item homogenitásvizsgálatát a WLE személyszeparációs reliabilitásmutatóval (PSR – *Person Separation Reliability*; Write és Stone, 1979) végeztük. [A PSR érték megmutatja, hogy a teszt itemei milyen mértékben képesek elkülöníteni egymástól a diákokat. Értéke a klasszikus tesztelméletben ismert reliabilitásmutatóhoz hasonlóan 0 és 1 közötti. Minél magasabb az érték, annál pontosabb a mérés, annál kisebb a mérés hibája. A teszt belső konzisztenciáját tekintve a PRS érték számolási és értelmezési módja párhuzamba

állítható a klasszikus tesztelméletben használt KR20 (*Kuder–Richardson*) formulával, ami dichotóm itemek használata esetén azonos a Cronbach- $\alpha$  értékkel. Mindkét eljárással a teszt belső konzisztenciája jellemezhető.] A vizsgált célpopuláció részére az itemek nehézségi szintjének megfelelőségét a diákok képességszintjét és az itemek nehézségi szintjét közös képességskálán megjelenítő személy/item-térkép segítségével állapítottuk meg. Az összehasonlító elemzéseket a közös képességskálára hozás után klasszikus tesztelméleti módszerekkel végeztük.

A fejlődést jellemző logisztikus görbe (*Molnár és Csapó, 2003*) paramétereinek meghatározása során csak azon évfolyamok eredményeit vettük figyelembe, ahol több mérési ponton is rendelkezünk adatokkal (3–8. és 11. évfolyam), illetve az adott évfolyamon és mérési ponton felvett teszt reliabilitásmutatója 0,70 felett volt. A görbeillesztés során négyparaméteres logisztikus görbe  $[F(x) = ((A-D)/(1+((x/C)^B))) + D]$ ; A: minimum aszimptota, B: meredekség, C: inflexiós pont, D: maximum aszimptota] függvényét használtuk.

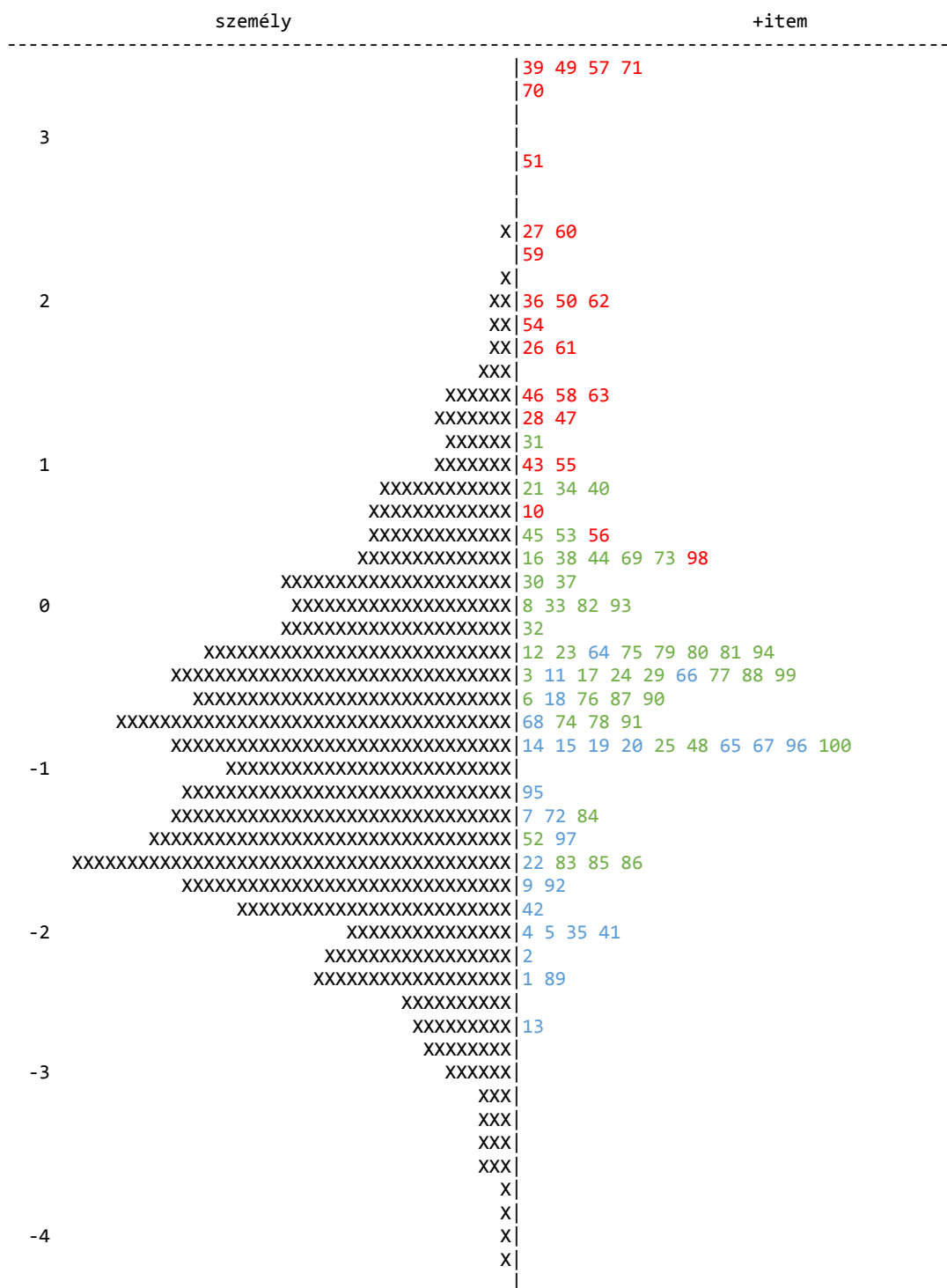
Az elemzésbe bevont tesztek reliabilitásmutatói mérésenkénti és évfolyamonkénti bontásban (Cronbach- $\alpha$ ) 0,70 és 0,86 közöttiek. A 3–11. évfolyam számára kidolgozott 100 item homogenitását jellemző WLE személyszeparációs reliabilitásmutató 0,80. A reliabilitásmutatók értékei alapján a kidolgozott és elemzésbe bevont itemek mind teszt, mind itembank szintjén alkalmasak a 3–11. évfolyamos diákok területspecifikus problémamegoldó képességének, a fejlettség szintjének jellemzésére.

A vizsgált populáció diákjainak képességszintjét és a közvetített 100 item nehézségi szintjét közös logitskálán fejezi ki a 4.6. ábra személy/item-térképe. A személy/item térkép alapján az elemzésbe bevont 100 item nehézség alapján is alkalmas a kiválasztott célpopuláció, azaz a 3–11. évfolyamos diákok területspecifikus problémamegoldó képességének, a fejlettségi szintek vizsgálatára. Teljesül az a feltétel, hogy a feladatbank itemeinek nehézségi szintje közel áll a diákok képességfejlettségi szintjéhez, azaz grafikus megjelenítésben a képesség/itemnehézségi logitskálán párhuzamosan helyezkednek el képességszint szerint a diákok és nehézségszint szerint az itemek. Az itembank továbbfejlesztése esetén a még pontosabb lefedés érdekében különös hangsúlyt kell fektetni az itembankban szereplő legkönnyebb itemeknél is könnyebb itemek kidolgozására.

Közös nehézségi skálán jellemezve az itemeket, jól azonosíthatóan elválnak nehézségi szint tekintetében egymástól a különböző típusú problémák. Legkönnyebbnek az iskolai tudás alkalmazását kívánó jóldefiniált problémák, átlagos nehézségűnek az iskolai tudás alkalmazását kívánó rosszuldefiniált problémák és legnehezebbnek az iskolán kívül tanultak alkalmazását kívánó rosszuldefiniált problémák bizonyultak.

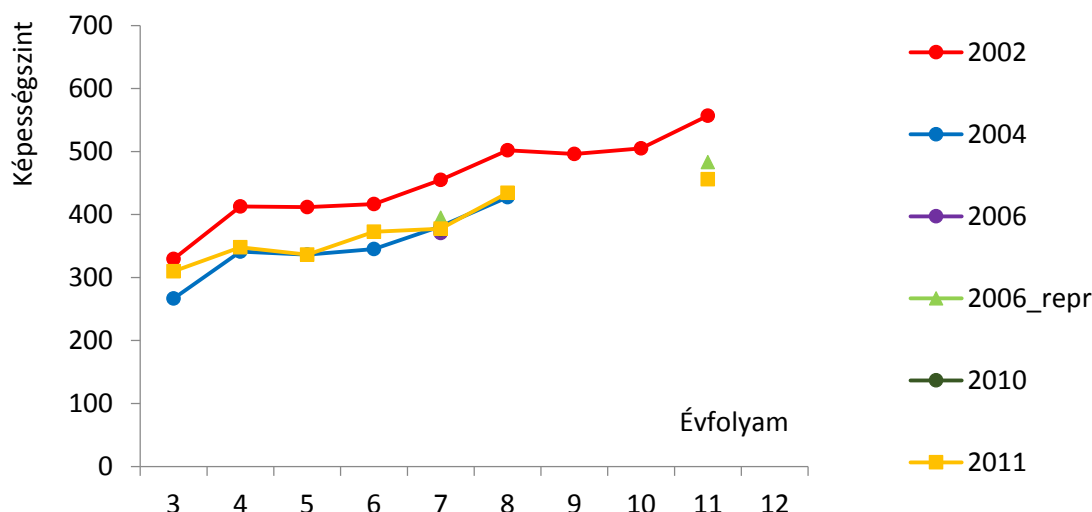
Az adatfelvétel időpontjától függetlenül a fejlődés ütemét jellemző görbék főbb tulajdonságaiban nem történt változás az elmúlt tíz év alatt (4.7. ábra). A fejlődés menetét jellemző görbék közel párhuzamosan futnak, meredekségük és inflexiós pontjuk tekintetében hasonló paraméterekkel rendelkeznek. A görbék egymáshoz való eltolódását egyrészt a minta tulajdonságai, másrészt a diákok átlagos képességszintjében történt változás okozza.

2002 és 2011 között minden évfolyamon átlagos képességszint-csökkenés figyelhető meg. Ez nem magyarázható az egyes adatfelvételek mintájának speciális tulajdonságaival, miután 2006 és 2011 között három reprezentatív, azonos tulajdonságokkal bíró mintán történt a mérés. Ezt támasztja alá a 7. évfolyamosok eredményének változása is: ezen diákok átlagos képességszintjéről minden egyes adatfelvétel kapcsán rendelkezünk adatokkal, amelyek a 2002 és 2011 között történt képességszint-csökkenésre utalnak.



4.6. ábra

A problémamegoldó képesség fejlettségi szintjének mérésére alkalmas feladatbank 100 itemének nehézségi indexe a 3–11. évfolyamos diákok képességszintjének függvényében (minden 'x' 39 diákot reprezentál; a színek a feladatok típusát mutatják; 1. típus – iskolai tudás alkalmazását kívánó jóldefiniált problémák: kék, 2. típus – iskolai tudás alkalmazását kívánó rosszuldefiniált problémák: zöld, 3. típus – iskolán kívül tanultak alkalmazását kívánó rosszuldefiniált problémák: piros)



4.7. ábra  
A problémamegoldó képesség fejlődése

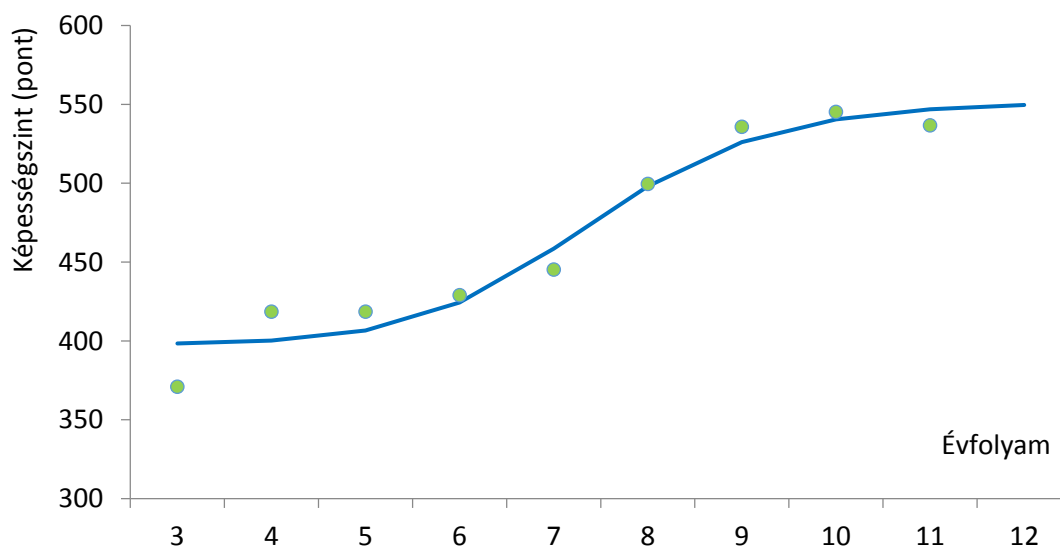
A 2002-es és a 2004-es eredmények közötti, átlagos, minden évfolyamot érintő jelentős mértékű képességszint-csökkenés bizonyos mértékben magyarázható a két kutatás mintájának főbb tulajdonságaival. A 2002-es minta nagyvárosi diákjaihoz képes 2004-ben hátrányos helyzetű diákok képességszintje került górcső alá, akik átlagosan 10-15 százalékkal teljesítettek alacsonyabban a teszteken, ami kétévnyi átlagos képességfejlődésnek felelt meg. Egy 8. évfolyamos hátrányos helyzetű diák problémamegoldó képességének fejlettségi szintje egy átlagos nagyvárosi 6. évfolyamos diák fejlettségi szintjének felelt meg. Miután a teljesítmények szórásában nem történt változás, közel azonos mértékű minden egyes évfolyamon és mindkét mintán, a fejlődés ütemének jellemzői nem a speciális (nagyvárosi vagy hátrányos) helyzetű diákok sajátja, hanem általános tendencia. A 2002-es és 2004-es adatfelvételek alapján felrajzolt fejlődési görbék párhuzamosak, ami arra utal, hogy az iskola nem csökkent a hátránnyal indulók helyzetén ezen a képességterületen, ugyanakkor az eredmények alapján azt nem is növeli tovább, ami mindenképpen pozitív jelenség (részletesebben l. *Molnár*, 2004).

A 2006-ban 7. évfolyamos diákok átlagos teljesítménye nem a nagyvárosi, hanem a hátrányos helyzetű diákok átlagos teljesítményéhez közeli, míg a 2010-es és a 2011-es szintén reprezentatív minta átlagos képességszintje 7. évfolyamon már a 2004-ben hátrányos helyzetű diákok átlagos képességszintjét sem érte el, annál szignifikánsan alacsonyabb volt. A 11. évfolyamos diákok esetében, ahol 2002-es, 2006-os és 2010-es adatokkal rendelkezünk, szintén hasonló tendenciát tapasztaltunk. A színvonalcsökkenés oka számos helyen kereshető: változás a tantervekben, az oktatás módszereiben, esetleg egyéb háttérváltozókban. Mindennek feltárása azonban további kutatást igényel.

A 2011-es mérés 3. évfolyamosokra vonatkozó eredménye a korábban említett tendencia megfordulását jelezheti. A 2011-ben 3. évfolyamosok átlagos képességszintje azonos a 2002-ben nagyvárosi 3. évfolyamos diákok átlagos képességszintjével ( $t=1,91$ ,  $p=0,06$ ). Ennek a hipotézisnek a hosszú távú igazolása további kutatást, az érintett diákok longitudinális követését igényli.

Összességében a fejlődés terén tapasztalt azonosságok és a kutatások felépítése lehetővé teszi az általános problémamegoldó gondolkodás képességskálájának kialakítását és a fejlődést leíró görbe jellemzését (4.8. ábra). A problémamegoldó képesség fejlődése a többi képesség fejlődési menetéhez hasonlóan jól jellemezhető logisztikus görbével. Az illesztett négyparaméteres görbe jól reprezentálja az empirikus adatokat, a determinációs együttható értéke ( $R^2=0,95$ ) jó. Minden évfolyamon stagnál vagy nő a diákok problémamegoldó képességének átlagos fejlettségi szintje, s a fejlődés mértéke, hasonlóan más általános gondolkodási képesség (pl. induktív gondolkodás) fejlődéséhez, viszonylag lassú. A 3–11. évfolyamok teljes időszakát tekintve évenként átlagosan a szórás ötödével fejlődik. Nem figyelhetünk meg fejlődést a 4. évfolyamon, míg a legintenzívebb fejlődés a 7. és 8. évfolyam között történik, közel kétszer annyi, mint az éves átlagos fejlődés mértéke. Ezt támasztja alá, hogy 7. évfolyamra esik a logisztikus görbe inflexiós pontja, azaz ebben az életkorban hatékonyan fejleszthető a diákok problémamegoldó képessége, majd ezen életkor után az addig gyorsuló fejlődés lassuló növekedésbe vált. Extrapolálva a fejlődés folyamatát, a 3. évfolyam előtt és a 11. évfolyam után is, bár lassuló ütemben, de folytatódik a képesség fejlődése.

Mindezek alapján alapvetően három szakaszra bonthatjuk a diákok problémamegoldó képességének fejlődését annak gyorsasága és fejleszthetősége szemszögéből: 6. évfolyamig egy lassú, de fokozatosan gyorsuló fejlődés figyelhető meg, amit a 7–8. évfolyamon egy intenzív fejlődési szakasz követ, majd a középiskolától ismét lelassul a fejlődés és ezzel párhuzamosan a fejleszthetőség mértéke is.

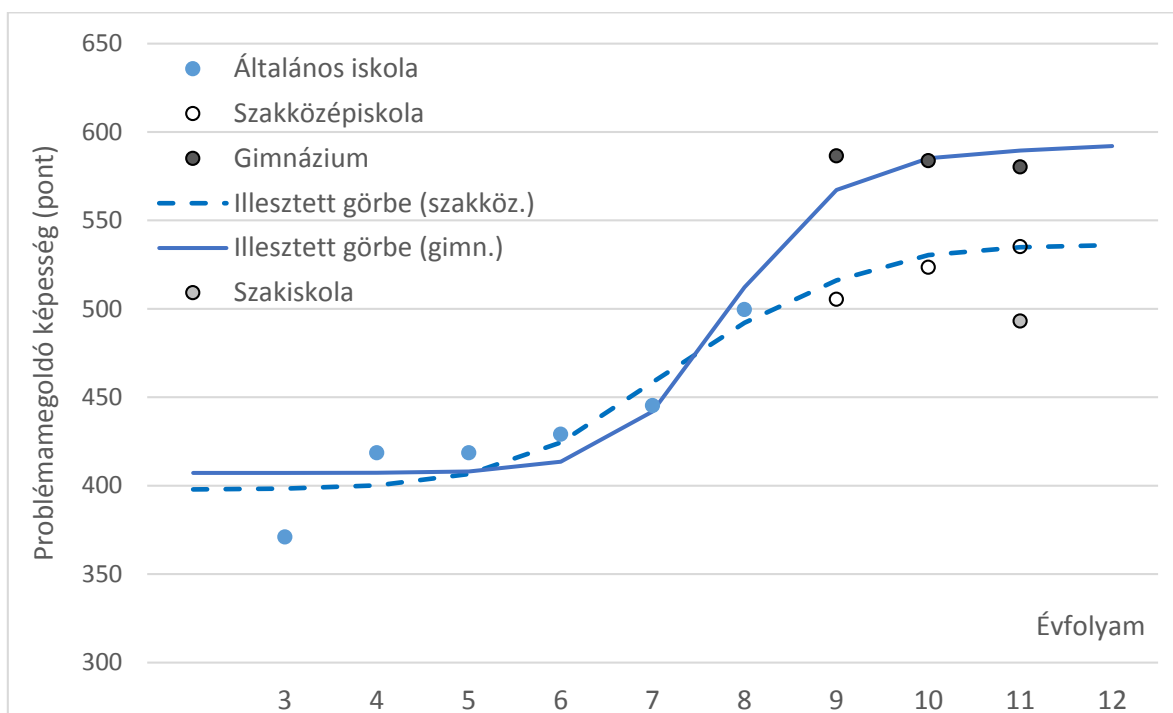


4.8. ábra

*A problémamegoldó képesség fejlődése 3–12. évfolyamon*

Jelentős mértékben változik a problémamegoldó képesség fejlődési trendvonalát jellemző görbe meredeksége és maximuma, ha 3–11. évfolyam vonatkozásában iskolatípusonkénti bontásban elemezzük az adatokat. Mind a gimnáziumba, mind a szakközépiskolába járó diákok átlagos fejlődése továbbra is jellemezhető és pontosan jellemezhető egy négyparaméteres logisztikus görbével. A determinációs együttható értéke mindkét esetben  $R^2=0,95$ , azonban a

két görbe inflexiós pontja továbbra is 14 év körül (7-8. osztály) mozog. Az általános iskola utáni szelekció jelentős mértékben felerősíti az azonos évfolyamon, de más iskolatípusban tanuló diákok közötti különbségeket (4.9. ábra). A különbség mértéke a középiskolai évek alatt enyhén csökken, de nagysága az éves átlagos fejlődés mértékének többszöröse: 9. évfolyamon megközelítőleg 80 pont, 11. évfolyamon 50 pont, miközben a 9. évfolyamos tanulók teljesítményének szórása 78 pont.



4.9. ábra

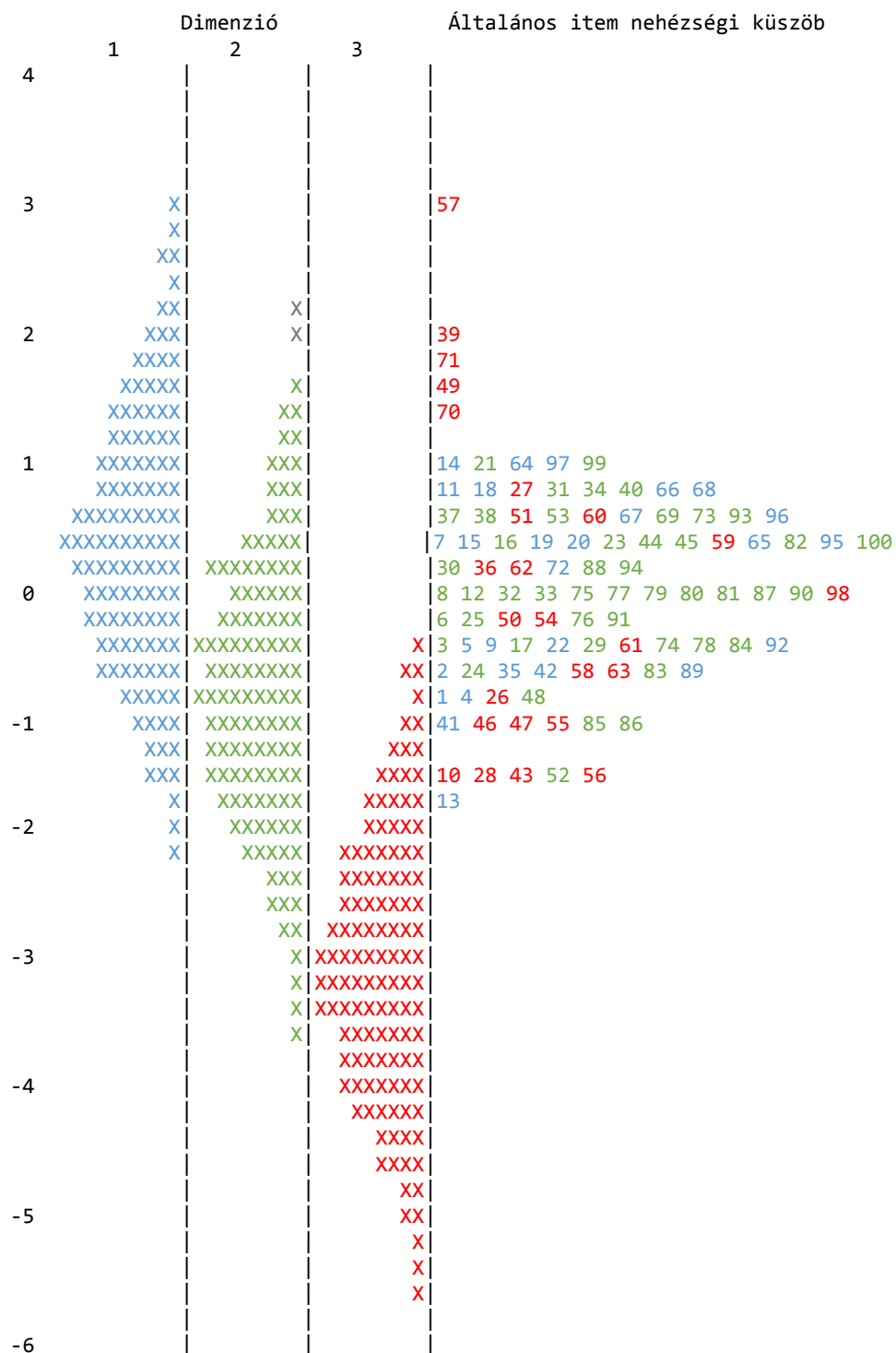
*A problémamegoldó képesség fejlődése iskolatípusonkénti bontásban*

A szelekció mértékét jellemzi, hogy egy 500 (100)-as skálán egyrészt 60 pont különbség van 10. évfolyamos gimnazisták és a szintén érettségit adó szakközépiskolás kortársaik átlagos teljesítménye között, miközben a gimnáziumban tanuló kortársaik átlagosan 600 pontot körüli teljesítményt mutatnak. A szakiskolás 11. évfolyamos diákok átlagos teljesítménye éri el a 8. évfolyamosok átlagos teljesítményét, több mint egy szórással alacsonyabban teljesítve gimnáziumba járó kortársaiknál. Extrapolálva a fejlődés folyamatát, megállapítható, hogy iskolatípustól függetlenül 11. évfolyam és a közoktatás vége után is, bár lassuló ütemben, de folytatódik e képesség fejlődése.

Az egyének szintjén még jelentősebbek a különbségek. A 3. évfolyamos diákok 5%-a egy átlagos 11. évfolyamos diák problémamegoldó gondolkodásának szintjén vagy a felett teljesít, míg a 11. évfolyamos diákok 15%-a az átlagos 3. évfolyamos képességszintet sem éri el.

A három különböző problématípus átlagos képességfejlettségi szintjében szintén jelentős mértékű, több évnyi átlagos fejlődésnek megfelelő szintű eltérés tapasztalható bármely mérési ponton. Az eltérés mértéke nagyobb, mint a 3–11. évfolyam vonatkozásában történt átlagos képességfejlődés nagysága. A különböző típusú problémákon nyújtott teljesítmények eloszlásgörbéinek logitskálán való elhelyezkedése (4.10. ábra) és a fejlődés ütemét leíró görbék

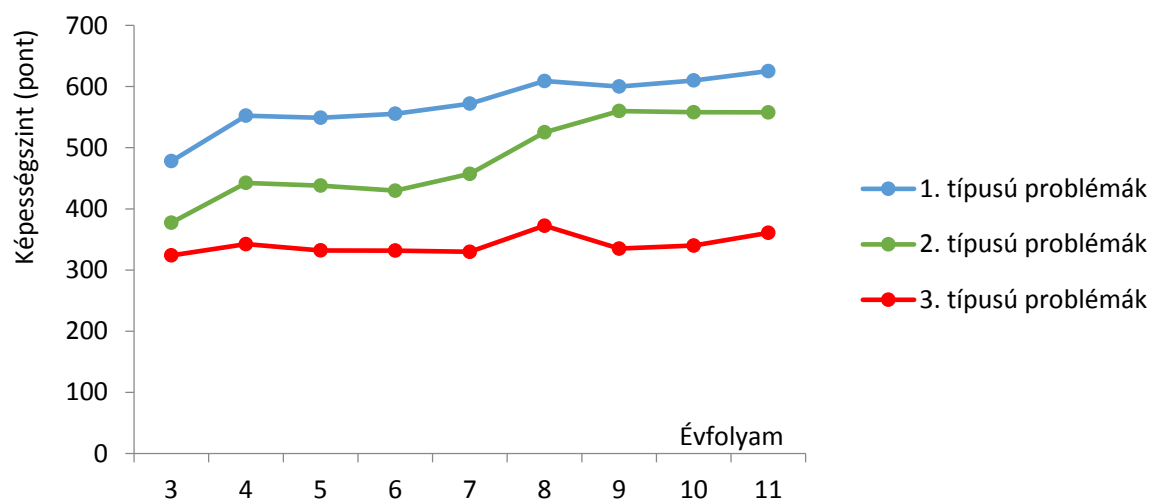
tulajdonságai (4.11. ábra) eltérőek. A teljesítmények eloszlását jellemző görbék mindhárom esetben normál eloszlásúak.



4.10. ábra

Különböző típusú problémák megoldásának személy/ item-térképe (1. dimenzió: iskolai tudás alkalmazását kívánó jóldefiniált problémák (kék); 2. dimenzió: iskolai tudás alkalmazását kívánó rosszuldefiniált problémák (zöld); 3. dimenzió: iskolán kívül tanultak alkalmazását kívánó rosszuldefiniált problémák (piros); minden 'x' 198 diákot reprezentál)





4.11. ábra

*Különböző típusú problémákon mutatott képességfejlődés (1. típus: iskolai tudás alkalmazását kívánó jóldefiniált problémák; 2. típus: iskolai tudás alkalmazását kívánó rosszuldefiniált problémák; 3. típus: iskolán kívül tanultak alkalmazását kívánó rosszuldefiniált problémák)*

A 3–11. évfolyam vonatkozásában kivétel nélkül azon problémák megoldásában bizonyultak legjobbnak a diákok, amelyek megoldáshoz minden szükséges információt tartalmazott a feladatlap, a rendelkezésre álló információk mellett nem vagy csak kevés zavaró információ fordult elő, és a problémák megoldása során főképp alapszempontokra kellett támaszkodni. E típusú problémák megoldásában már a 3. évfolyamos diákok képességi szintje (478 pont) a mindhárom típusú probléma megoldása alapján becsült 7. évfolyamos átlagos (445 pont) képességi szint felett van. A 8. és a 11. évfolyamos diákok átlagos képességi szintje a legmagasabb képességtartományban, az átlag felett több mint egy szórásnyi terjedelemben helyezkedik el. A legjelentősebb fejlődés e típusú problémák megoldásában, előzetes hipotézisünknek megfelelően, általános iskolában (131 pont), azon belül a 3–11. évfolyam vonatkozásában a 3. és 7. évfolyamon következik be. Ugyanakkor a várttal ellentétben még középiskolában is kimutatható az alapszempontokra támaszkodó problémák megoldásán mutatott fejlődés, igaz, mértéke az általános iskolában tapasztalt átlagos fejlődés felének felel meg (16 pont).

A második típusba sorolt problémák megoldásában tapasztalható mind általános (180 pont), mind középiskolában (32 pont) a legjelentősebb fejlődés. Ezen problémák megoldásához nem minden információt adtunk meg, ugyanakkor a problémák megoldásához szükséges információk a tananyag részét képezik. A képességfejlődés üteme hatodik évfolyamig párhuzamos az első típusú problémák esetén tapasztaltakkal, majd az átlagos fejlődés ütemének nagyobb mértékű gyorsulása összességében jelentősebb képességfejlődést eredményez.

Az általános iskolában tapasztalt fejlődés mértéke azonos az első típusú problémák megoldása esetén általános és középiskolában összességében tapasztalt fejlődés mértékével (147 pont). A fejlődés mértéke e típusú problémák esetén is lelassul 8. évfolyam után.

A teljesítményszintek jelentősebb mértékű csökkenésének és a fejlődési görbe markáns változásának lehetünk tanúi a harmadik típusú problémáknál. E hiányos, ugyanakkor szemantikailag gazdag problémák hasonlítanak leginkább a valós élet problémáihoz, ahol a



megoldáshoz szükséges információk nem állnak teljes mértékben rendelkezésre, sőt az információk között – nagy valószínűséggel – van olyan is, amelyekkel a diákok nem találkoztak az iskolában. E problémák megoldása tükrözi leginkább a diákok valós helyzetekben történő problémamegoldásának sikerességét. A teljesítmények eloszlásgörbéje ebben az esetben is normál eloszlást mutat, azonban a görbe képességskálán való elhelyezkedése arra utal, hogy az e típusú problémák legjobb problémamegoldói érik csak el azt a képességfejlettségi szintet, ami a mintában egy átlagos, második típusú problémákat megoldó diákokat jellemez.

A képességfejlődést leíró görbe alapján a diákok 7. évfolyamig nem mutatnak jelentős mértékű fejlődést, majd a 7. évfolyamon tapasztalt fejlődés után, az előzetes hipotézisünkkel ellentétben, középiskolában sem tapasztalható jelentősebb mértékű fejlődés e problémák megoldása terén. A 3–11. évfolyam kumulált 37 pontos fejlődése arra utal, hogy az e típusú problémák megoldását nem fejleszti az iskola, holott a mindennapi problémák jelentős része e problémaosztályba sorolható.

A nem teljesítménybefolyásoló szerepe nem változott az elmúlt tíz év során, a fiúk és a lányok teljesítménye között, hasonlóan más képességterületen tapasztaltakhoz, nem mutatható ki szignifikáns különbség. Mind 2002-ben, mind 2006-ban és 2011-ben a középiskola vége felé, 11. évfolyamon jelentkezik egyértelműen ( $p < 0,01$ ) a fiúk előnye és átlagosan magasabb képességszintje.

A szülő iskolai végzettségének teljesítményt meghatározó szerepe változott az elmúlt évek alatt. Általános tendencia, hogy a szülők átlagos iskolai végzettsége fél-egy fokozatot nöött 2002 és 2011 között. Míg 2002-ben átlagosan szakiskolai végzettséggel, addig 2011-ben már érettségivel rendelkezik egy átlagosnak nevezhető szülő. Azonban mindez nem vonta maga után a diákok problémamegoldó képességszintjének átlagos növekedését. Az adatokból egy új tendenciára következtethetünk: míg 2002-ben általános iskolában gyenge ( $r=0,1-0,2$ ), középiskolában egyre erősödő ( $r=0,3-0,4$ ) kapcsolat volt megfigyelhető a szülő iskolai végzettsége és a diákok teszten mutatott teljesítménye között, addig 2011-re ez az összefüggés megszűnt. A 3. és az 5. évfolyam kivételével, ahol a gyenge ( $r=0,1$ ) kapcsolat még kimutatható, nincs szignifikáns összefüggés sem általános, sem középiskolában a diákok mért teljesítménye és a szülők iskolai végzettsége között.

#### 4.1.2.1. A területspecifikus problémamegoldó képesség papíralapú mérését célzó kutatások szintetizálása – konklúzió

Felépítettünk egy 100 feladatból álló, a 3–11. évfolyam vonatkozásában hatékonyan működő és használható, megfelelő mutatókkal rendelkező, a területspecifikus problémamegoldó képesség fejlettségét vizsgáló papíralapú itembankot. A feladatbank felhasználásával közös képességskálán jellemeztük a 3–11. évfolyamos diákok problémamegoldó képességének fejlettségi szintjét és annak változását közel tíz év távlatában. Miután a feladatbankban különböző típusú problémák szerepeltek [(1) alpműveletek alkalmazásával megoldható problémák, amelyek megoldásához minden információt tartalmazott a probléma; (2) problémák, amelyek megoldásához nem minden információt tartalmazott a probléma, de azok a tananyag részét képezik; illetve (3) problémák, amelyek megoldásához szintén nem adtunk meg minden szükséges háttérinformációt, de azokkal a diákok, ha nem is az iskolai tanórán, de a hétköznapi életben találkozhatnak] lehetőségünk nyílt e három, egymástól jól elkülöníthető tulajdonságokkal bíró problémák megoldása során

mutatott fejlődésbeli különbségek és azonosságok leírására. Végül kiemeltük a szülő iskolai végzettsége teljesítménybefolyásoló hatásában történt jelentős mértékű változását e képességterületen.

2002 és 2011 között minden évfolyamon átlagos képességszint-csökkenés figyelhető meg, ami az eredmények alapján nem magyarázható kizárólagosan az egyes adatfelvételek mintájának speciális tulajdonságaival. A 2011-es adatfelvétel 3. évfolyamos diákokra vonatkozó eredményei e tendencia megváltozását jelzik, ugyanakkor ez további kutatást igényel. Összességében az oktatás nem csökkenti, ám nem is növeli tovább a hátránnyal indulók helyzetét ezen a képességterületen.

A problémamegoldó képesség fejlődése is – hasonlóan más gondolkodási képesség fejlődéséhez – jól jellemezhető egy négyparaméteres logisztikus görbével. A görbe inflexiós pontja, mérési ponttól függetlenül, 7. évfolyamra esik, azaz ebben az életkorban a leggyorsabb a fejlődés és ezért hatékony a fejleszthetőség, majd ezen életkor után az addig gyorsuló fejlődés lassuló növekedésbe vált. A fejlődés mértéke – hasonlóan más általános gondolkodási képesség fejlődéséhez lassú, 3–11. évfolyamok teljes időszakát tekintve évenként átlagosan a szórás ötöde. 4. évfolyamon minimális, 7-8. évfolyamon intenzív fejlődés figyelhető meg, míg extrapolálva a fejlődési folyamatokat a 3. évfolyam előtt és a 11. évfolyam után is, bár lassuló ütemben, de folytatódik e képesség fejlődése. Több éves fejlődésnek megfelelő mértékű teljesítménykülönbség van a különböző iskolatípusban tanuló diákok között, ami egy évfolyamon belül a diákok szintjén még markánsabban realizálódik.

A három különböző problématípus megoldási sikeressége alapján megállapított képességszintek közötti különbség mértéke egy évfolyamon belül is jelentős, években mérhető. A valós helyzetekben történő problémamegoldás sikerességét leginkább tükröző rosszul definiált, hiányos és szemantikailag gazdag problémahelyzetek évfolyamtól függetlenül jelentős kihívást jelentettek a diákoknak. Képességfejlődésük alacsony mértéke arra utal, hogy e típusú problémahelyzetek megoldására egyáltalán nem készíti őket fel az iskola és spontán fejlődése igen lassú.

A szülő iskolai végzettségének korábban jelentős mértékű teljesítménybefolyásoló szerepe változott az elmúlt tíz év alatt. Egyrészt általános tendencia, hogy az oktatás expanziójával a szülők átlagos iskolai végzettsége is növekedett, ez azonban – a várakozásokkal ellentétben – a problémamegoldó gondolkodás képességterületén nem vonta maga után a diákok átlagos képességszintjének növekedését. Másrészt új tendencia, hogy a korábban az egyik legjelentősebb előrejelző és képességszint-befolyásoló tényezőnek számító anya iskolai végzettsége teljes mértékben megszűnőben van, 2011-ben egyik évfolyamon sem volt kimutatható közepes vagy erős szignifikáns kapcsolat a diákok problémamegoldó képességének fejlettségi szintje és az anyák iskolai végzettsége között.

Az utóbbi években tapasztalt jelentős mértékű fejlődés a pedagógiai mérés-értékelés terén, valamint a technológia rapid változása, rohamos elterjedése és iskolai megjelenése új lehetőségeket teremtett a problémamegoldó képesség vizsgálata terén, háttérbe szorítva a korábbi papíralapú, statikus problémahelyzetekre alapozó kutatásokat, és előtérbe állítva a 21. századra jellemző dinamikusan változó, interaktív problémahelyzetek segítségével megvalósuló méréseket. A technológia adatfelvételbe történő integrációja olyan új eszközöket (pl.: multimédia alkalmazása, dinamikusan változó problémakörnyezet, társas problémamegoldás) kínált és kínál a problémamegoldó képesség fejlődését és fejlesztési

lehetőségeit vizsgáló kutatók számára, melyek forradalmasítják a korábbi problémamegoldásra vonatkozó modelleket és teszteket.

Az átállás első fázisában a problémamegoldó képesség mérésével kapcsolatos hazai kutatásainkra is – a többi képességtérületen alkalmazott eljáráshoz hasonlóan – a fokozatosság jellemző. Első lépésben még nem törekedtünk dinamikus elemek megjelenítésére, a hangsúly a közvetítő eszköz befolyásoló hatásának meghatározásán volt. A vonatkozó adatfelvételt 2009 tavaszán 6. évfolyamos diákok részvételével (n=598) végeztük. A minta egyik része papír (PP), másik része számítógép-alapon (CB) oldotta meg a korábbi papíralapú kutatásokban is alkalmazott 28 ítemes problémamegoldó feladatlapot. Az online adatfelvétel a TAO platform segítségével történt. A kontextus és az alkalmazott itemtípusok következtében a papír és a számítógép-alapú feladatlap nem volt teljesen azonos, egyes problémák megoldása során más-más típusú itemet alkalmaztunk a két médium esetében. A papíralapon feleletalkotó típusú kérdéseket feleletválasztó típusú kérdésekkel helyettesítettük online környezetben. Az adatfelvétel során megtartottuk a teszt linearitását, azaz minden egyes diák ugyanabban a sorrendben kapta ugyanazokat a feladatokat. Az alkalmazott online platform azonban nem tette lehetővé, hogy a papíralapú tesztfüzet egy oldalán található feladatok az online adatfelvétel során egyszerre jelenjenek meg a képernyőn. A tesztfeladatok közötti navigálást egy előre és visszafelé lépést lehetővé tevő navigációs gombbal valósítottuk meg. A feleletválasztós kérdésekre adott választ papíron a diákok karikázással adták meg, számítógépen rádiógomb használatával, kattintással kellett kiválasztatni a helyesnek ítélt megoldást. A feladatlap megoldására mindkét esetben egy tanítási óra állt a diákok rendelkezésére. Az elemzés során mind klasszikus, mind valószínűségi tesztelméleti eszközöket használtuk. A kutatás eredményeiről részletesen l. *Molnár* (2010a, 2010b) tanulmányát. A problémamegoldó képesség mérésére alkalmas papíralapú teszt egyszerű lineáris formában történő digitalizálása esetén is megmutatkozott a számítógép-alapú tesztelés egyik előnye, színesek és ezáltal élvezhetőbbek, motiválóbbak lettek a megoldandó problémák. A teszt azonban továbbra is első generációs, hagyományos, lineáris teszt maradt, és nem használta ki a multimédia adta lehetőségek széles skáláját.

A számítógépes tesztelésre való átállás következő lépcsőfokán második generációs tesztek (*Pachler* és mtsai, 2010) alkalmazására került sor, azaz bár ismételten statikus, de már a multimédia adta lehetőségeket kihasználó (pl. hangot tartalmazó) problémahelyzetek kiközvetítése történt meg. Ezzel a megoldással az olvasni még nem tudó diákok is részt tudtak venni a tesztelésben, miután a megoldandó problémák gombnyomásra meghallgathatóak voltak. Ezzel az eljárással jelentős mértékben csökkenthető a problémák megismeréséhez szükséges olvasási képesség fejlettségi szintjének teljesítményt meghatározó ereje.

A számítógép-alapú egyéni tesztelés ma adott lehetőségeit teljes mértékben kihasználva a papíralapú tesztelésről való eltávolodás harmadik fázisában mind a multimédia adta lehetőségek (pl.: a megoldandó problémák meghallgathatóak), mind az interaktivitást biztosító dinamikusan változó elemek megjelentek a problémákban. Ezzel megtörtént a harmadik generációs tesztek adta lehetőségek kihasználása, amelyek már sem a hagyományos papíralapú, sem az első és második generációs számítógép-alapú teszteléssel nem érhetőek el. Harmadik generációs tesztek alkalmazásával lehetővé vált a problémamegoldó képesség korábban nem vizsgált aspektusainak kutatása (l. *Molnár*, 2012; *Greiff, Wüstenberg, Molnár, Fischer, Funke* és *Csapó*, 2013; *Wüstenberg, Greiff, Molnár és Funke*, 2014).

#### 4.2. A problémamegoldó képesség számítógép-alapú mérése: célok és módszerek

A 21. század polgára már nem boldogulhat az életét minden területen körbevevő interaktív technológiai eszközök használata nélkül. Ezek meghatározzák szórakozási tevékenységeinket, munkánkat, kommunikációs szokásainkat. Az mp3 lejátszó, a távirányító, a televízió, a mobiltelefon, a fénymásoló, sőt ma már a mosógép beindításához és használatához is az adott géppel történő interakciók sorozatára van szükség: különböző gombok bizonyos sorrendben történő lenyomására, tekerésére, esetleg az eszközt vezérlő program érintőképernyőn keresztül történő irányítására. A mindezt lehetővé tevő új szoftveres és hardveres technológiák állandó tanulásra és ezzel párhuzamosan problémamegoldásra készítetnek minket. Ma már természetesnek vesszük, hogy mindenki tudja kezelni automatizált környezetét, mobiltelefonját, háztartási eszközeit, autóját, a munkahelyén lévő technikai eszközöket. A 21. század embere egy nap alatt a korábban nem tapasztalt mennyiségű interakciót folytat le a különböző technológiai eszközökkel, miközben nem egyedül, hanem csoportban, mások ismereteit, problémamegoldó stratégiáit integrálva, véleményét, személyiségét kezelve dolgozik. De hogyan lehet ezeket a 21. században kulcsfontosságúnak számító képességeket mérni? Lehetséges-e ezen képességek fejlettségi szintjének meghatározása, előrejelzése hagyományosnak számító első generációs teszteredmények alapján? Hogyan fejlődik a diákok e típusú, interaktív környezetben történő problémamegoldó képessége? A fejlődés menete illeszkedik-e a korábbi problémamegoldó, illetve gondolkodási képességeket vizsgáló kutatási eredményekhez? Készen állnak a diákok arra, hogy harmadik generációs tesztekkel mérjük képességszintjeiket, vagy inkább az első és második generációs tesztek preferálják?

Az iskolai nevelés és oktatás egyik fontos feladata az értelmi képességek fejlesztése (*Gordon Győri*, 2002), mely során lényeges feladat a fejlesztésre érzékeny szakasz meghatározása (l. pl.: 4. és 5. fejezet), ezt követően az értelmi képességek fejlődési tendenciáinak ismertetése. A problémamegoldó képességet és a képesség fejlődését ezért számos képesség és egyéb háttérváltozó fényében vizsgálták már. Miután már az intelligencia kezdeti definícióiban is helyet kapott a különböző problémák megoldási képessége, valamint az induktív gondolkodás az új tudás megszerzésének alapvető képessége, ezért lényeges a három konstruktum (problémamegoldó képesség, intelligencia és induktív gondolkodás) egymáshoz való viszonyának meghatározása, illetve az ezek fejlettségi szintjét befolyásoló tényezők azonosítása.

A problémamegoldó gondolkodás és az intelligencia kapcsolatának meghatározása nem új keletű probléma. Már a kezdeti kutatásokban megfigyelhető a problémamegoldó gondolkodás és az intelligencia kapcsolatának kettős megközelítése. A kutatások egy része független képességnek feltételezte a problémamegoldó képességet és az intelligenciát. Ezt számos empirikus kutatás alátámasztotta [Raven Standard Progresszív Mátrix (*Raven*, 1962) kapcsán *Beckmann* és *Guthke*, 1995 vagy *Revákné Markóczy*, 2001; az általános g-faktor vonatkozásában *Danner* és *mtsai*, 2011], míg több kutatás a problémamegoldó gondolkodás különböző dimenziója kapcsán cáfolta ( $r=0,65$  – *Kröner*, *Plass* és *Leutner*, 2005;  $r=0,34$  és  $r=0,43$  – *Rigas*, *Carling* és *Brehmer*, 2002;  $r=0,33$  és  $0,65$  – *Gonzalez*, *Thomas* és *Vanyukov*, 2005;  $r=0,56$  és  $0,63$  – *Wüstenberg*, *Greiff* és *Funke*, 2012). Utóbbiak általánosságban közepes kapcsolatot bizonyítottak az intelligencia (általános g-faktor) és a problémamegoldó képesség fejlettségi szintje között.

Ugyanakkor a kezdeti kutatások másik része az intelligencia eredeti meghatározása alapján azt a problémamegoldó képesség fejlettségi szintje legjobb előrejelzőjének tekintette (Putz-Osterloh, 1981; Putz-Osterloh és Lüer, 1981), jelentős mértékű oksági kapcsolatot feltételezve a két képesség között (Wenke, Fresch és Funke, 2005).

Azonban a számos releváns kutatás ellenére összességében megfogalmazható a kutatások hiányossága, hogy azok közel kizárólagosan laboratóriumi környezetben és egy kohorszra fókuszálva vizsgálták a diákok problémamegoldó képességének és intelligenciájának fejlettségi szintjét annak ellenére, hogy a készségek, a képességek tág életkori spektrumban, a fejlődés dimenziójából is vizsgálhatók. Az elemzések alapját képező kutatásban mindkettőt iskolai kontextusban és széles életkori intervallumban tettük.

Az induktív gondolkodás és a problémamegoldó képesség fejlettségi szintjének kapcsolatára is több kutatás fókuszált már, azonban azok nem különböztették meg egymástól a statikus és a dinamikus problémakörnyezetben nyújtott problémamegoldó sikerességet (l. pl.: Sternberg, 1994). E kutatások eredményei szerint (pl.: Chi, Glaser és Rees, 1982; Johnson-Laird, 1983; Klauer, 1989, 1996, 1992; Polya, 1954) az induktív gondolkodás fejlettségi szintje nagy jelentőséggel bír problémamegoldó képességünk (pl.: hipotézis felállítása és tesztelése; Gilhooly, 1982) fejlettségi szintje tekintetében.

A problémamegoldó képesség korai információ-feldolgozó modelljei központi szerepet tulajdonítottak a szabály indukciónak, ami az induktív gondolkodás alapeleme (Egan és Greeno, 1974; Simon, 1974; de Konig, 2000). A jelenkor kognitív nézetei értelmében a probléma megoldása során történő információfeldolgozás folyamatában alkalmazzuk leginkább az induktív gondolkodás képességét (Mayer, 1998), ami ezért hatással van a probléma megoldása közben történő tudás elsajátítás és tudás alkalmazás sikerességére (Bisanz, Bisanz és Korpan, 1994; Pellegrino és Glaser, 1982; Klauer, 1990, 1996, 1997; Hamers, De Koning és Sijtsma, 2000). Néhány kutatás (lásd pl.: Klieme, 2004; Wirth és Klieme, 2004) javasolja a terület-specifikus problémák megoldása során működtetett képességek elkülönítését a terület általános, komplex, tudás elsajátítására építő problémák (OECD, 2010c) megoldása során alkalmazott képességektől. Összességében kevés olyan kutatás ismert, ami megkülönbözteti a különböző problémahelyzetekben nyújtott teljesítmény alapján megállapított problémamegoldó képességszintet és azt kapcsolatba állítja a tesztelt személy induktív gondolkodásának fejlettségi szintjével (Molnár, Greiff, Wüstenberg és Fischer, 2016; Molnár, Greiff és Csapó, 2013).

A fejezet további részében szintetizáljuk a dinamikus problémamegoldó képesség mérésével kapcsolatos elemzések céljait, a kutatások módszereit (mintáját, mérőeszközeit, eljárásait). Részletesen kitérünk a harmadik generációs dinamikus problémamegoldó képességteszt működésére és hatékonyságára (reliabilitás, faktorstruktúra, mérési invariancia és a problémák skálázása szempontjából), valamint a tesztben lévő problémák különböző szcenáriók mentén történő viselkedésének alakulására.

#### 4.2.1. Célok

Az oktatási kontextusban alkalmazásra kerülő tesztek eredményeinek pontossága és validitása egyrészt azok elméleti alátámasztottságán, másrészt a tesztek jó pszichometriai jellemzőin múlik (American Educational Research Association American Psychological Association és National Council on Measurement in Education, 1999; Benson, Hulac és

Kranzler, 2010). Azaz, ha a tanulók interaktív problémamegoldó képességének fejlettségi szintje áll egy kutatás fókuszában, első lépésként szükséges egy minden szempont szerint megfelelő mérőeszköz kidolgozása. A nemzetközi együttműködés keretein belül fejlesztett mérőeszközsorozat minimálisan komplex rendszerekre épít, mely a problémamegoldás két fázisát a problémamegoldó képesség működését jellemző próba és hiba (*trial and error*) stratégia alkalmazásával történő ismeretek elsajátításának és ismeretek alkalmazásának sikerességét méri. A disszertációban bemutatásra kerülő elemzések elsődleges célja

- egy tág életkori intervallumban alkalmazható mérőeszközsorozat összeállítása;
- a mérőeszközök pszichometria jóságmutatóinak jellemzése mellett a mérőeszközfejlesztés mögött lévő elméleti kétdimenziós mérési modell tesztelése;
- a különböző csoportok összehasonlítását lehetővé tevő mérési invariancia tesztelése, annak ellenőrzése és számszerűsítése, hogy a teszt ugyanazon felépítésű konstruktumot mér-e (Byrne és Stewart, 2006) a különböző évfolyamokon, azaz releváns-e a teszten nyújtott teljesítmények egymással történő összehasonlítása ezen csoportok között;
- annak számszerűsítése, hogy a problémák megoldása során adott feltételrendszer (pl.: a modellépítés során történő korlátozások, a probléma megoldásához rendelkezésre álló idő változása, a probléma meghallgathatósága) mennyiben befolyásolja a problémák nehézségi indexét és a teszteredményeket;
- a különböző évfolyamos és iskolatípusban tanuló diákok teljesítményének összehasonlítása, fejlődési trendvonalak tulajdonságainak azonosítása és összevetése korábbi problémamegoldó képesség fejlődésére vonatkozó eredményekkel, illetve más, gondolkodási képességeket vizsgáló (pl.: induktív gondolkodás) kutatási eredményekkel;
- az iskoláztatás éve alatt a problémamegoldó képesség fejlesztésére érzékeny időszak meghatározására;
- a problémamegoldó képesség, az intelligencia, az induktív gondolkodás és a diákok tanulmányi sikeressége (iskolai jegyek) és gazdasági-társadalmi, szociális háttérváltozói (pl. szülők iskolai végzettsége) összefüggéseinek, strukturális relációinak és azok időbeli állandóságának vagy esetleges változásának feltérképezése, konstruktum validitás tesztelése;
- annak számszerűsítése, hogy lehetséges-e harmadik generációs tesztekkel mérhető képességek fejlettségi szintjének meghatározása, előrejelzése hagyományosnak számító első generációs teszteredmények alapján;
- a metaadatok segítségével végzett logfájl elemzések alapján van-e összefüggés a problémamegoldásra fordított idő és a problémamegoldás sikeressége között;
- a logfájl elemzések alapján milyen problémamegoldó stratégiák, milyen típusú profillal jellemezhető problémamegoldók azonosíthatóak;
- a metaadatok szerint hogyan változik a diákok által alkalmazott problémamegoldó stratégia időben a dinamikus problémamegoldó teszt megoldása közben.

## 4.2.2. Módszerek

## 4.2.2.1. A kutatások mintája

A kutatások mintájának főbb tulajdonságait a 4.2. táblázat foglalja össze. Az első harmadik generációs tesztek használó hazai kutatásunk 2011-ben volt 3–9. évfolyamos diákok bevonásával (N=855).

4.2. táblázat. A dinamikus problémamegoldó képességet mérő tesztekkel történő kutatások mintáinak főbb tulajdonságai

Adatfelvétel éve	Évfolyam	N	Főbb tulajdonság
2011	5.	26	n>5000 fős mintából véletlenszerűen generált (diákszintű mintaválasztás)
	6.	190	
	7.	157	
	8.	141	
	9.	83	
	10.	138	
2013	11.	120	pilot kutatás (osztályszintű mintaválasztás)
	3.	203	
	4.	260	
	5.	237	
	6.	170	
	7.	225	
	8.	123	
2014	9.	73	országos reprezentatív (longitudinális kutatás; osztályszintű mintaválasztás)
	7.	2237	
2014	3.	584	osztályszintű mintaválasztás
	4.	679	
	5.	608	
	6.	677	
	7.	607	
	8.	942	
	9.	30	
	10.	84	
	11.	102	
	12.	58	
2015	egyetem, első évfolyam	1259	egyetemszintű, önkéntes
2011–2014 együtt	3-12 + 1. évfolyam	8967	-

A kutatás egy része volt a 2011-es korábban bemutatott empirikus adatfelvételnek, aminek keretein belül összességében n>5000 diák oldott meg különböző kutatási elrendezés alkalmazása mellett papíralapú első generációs problémamegoldó képességet mérő tesztet, induktív gondolkodás tesztet, intelligenciatesztet, valamint a jelen fejezet fókuszában álló

számítógép-alapú problémamegoldó tesztet, illetve kitöltött egy háttérkérdőívet. Ezt követte 2013-ban egy 3–9. évfolyamos mintán (N=1291) történő adatfelvétel, mely alapvetően más kutatási elrendezést követett felhasználva a 2011-es adatfelvétel tapasztalatait és eredményeit.

A 2014-es adatfelvételek célja a felmerült kutatási kérdések pontosabb megválaszolását lehetővé tevő kutatások kivitelezése. Egyrészt a problémamegoldó képesség fejlettségi szintjét az iskolába lépéskor esetlegesen már prediktáló változó azonosítása, másrészt egy tág életkori intervallumban működő, harmadik generációs feladatokból álló feladatbank összeállítása.

A 2014-es első kutatás az MTA-SZTE Képességfejlődés Kutatócsoport Szegedi Iskolai Longitudinális Program keretein belül történt (N=2237), ami lehetőséget biztosít az interaktív környezetben történő problémamegoldó gondolkodás fejlettségi szintjét esetlegesen prediktív változók azonosítására. A disszertáció 5.3.3. fejezetében ezen adatfelvétel eredményei alapján a diákok első évfolyamos korában, különböző területeken nyújtott teljesítményének előrejelző hatását elemeztük hetedik évfolyamos problémamegoldó képességük fejlettségi szintje vonatkozásában. A longitudinális mintában 2013/2014-es tanévben hetedik évfolyamos diákoknál még első évfolyamos korukban (2007/2008-as tanévben) az iskolába lépés után felvételre került a DIFER tesztek egy speciálisan, a longitudinális mintában alkalmazott része, majd a májusi-júniusi adatfelvételi periódusban ugyanezen diákok megoldottak egy-egy szövegértés, egy számolási készség és egy induktív gondolkodás fejlettségi szintjét vizsgáló tesztet. Mindegyik teszt megoldására egy teljes tanítási óra állt rendelkezésükre. A tesztek 2007/2008-ban papír alapon kerültek kiközvetítésre. A 2014-ben megvalósított második kutatás felépítése és a valószínűségi tesztelmélet eszközeinek alkalmazása segítségével összeállítható egy tág életkori intervallumban jól mérő, adaptív tesztelésre is alkalmas feladatbank.

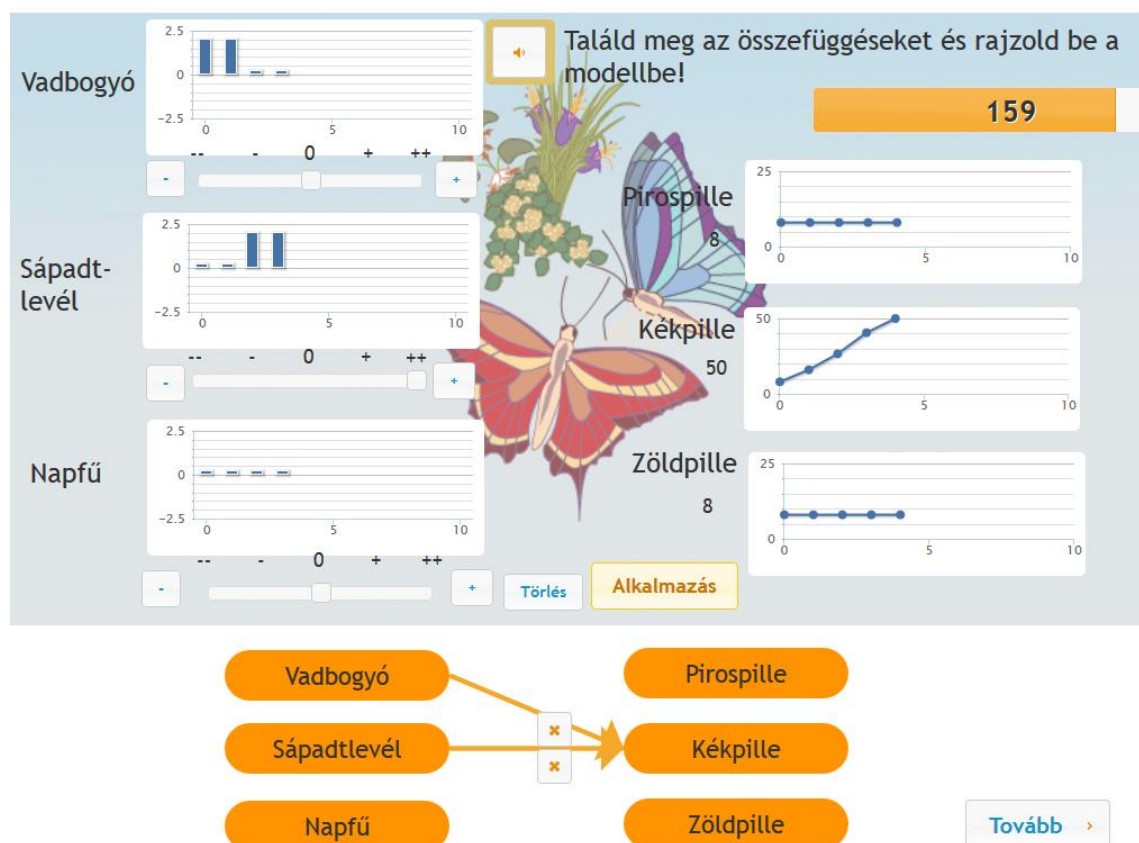
2015 szeptemberében a Szegedi Tudományegyetemet kezdő, 2015-ben érettségizett diákok úgynevezett bemeneti mérése keretein belül több teszt (matematika, magyar, történelem, angol, természettudomány, tanulási stratégiákra, demográfiai adatokra és érettségi eredményekre kérdező háttérkérdőív) mellett felvételre került a dinamikus problémamegoldó képesség fejlettségi szintjét vizsgáló mérőeszköz is. E kutatás adatbázisa lehetőséget teremt az általános és középiskolás diákok körében végzett kutatási eredmények kiterjesztésére és pontosítására.

#### 4.2.2.2. Méréseszközök és eljárások

A kutatásokban alkalmazott problémák a MicroDYN megközelítésen alapultak (részletes ismertetést l. 4.3. részt), és felépítésükben azonosak voltak a PISA 2012 kreatív problémamegoldás modul kutatásban alkalmazott interaktív problémákkal (OECD, 2014). A számítógép-alapú problémák a diákok által kedvelt, ismerős (pl.: mindennapi élet, videojátékok) kontextusban kerültek megfogalmazásra, ugyanakkor szerkezetük miatt számukra újak voltak, a megoldás során előzetes ismereteiket nem tudták alkalmazni. A teszt kiközvetítése az első kutatás keretein belül a TAO platformon, a második, harmadik és negyedik adatfelvétel során az eDia platform segítségével történt (a két platform feladatmegjelenítésében voltak kisebb különbségek, v.ö. 4.12. és 4.13. ábra). Mind a négy adatfelvételre az érintett iskola számítógépes termeiben került sor. A probléma megoldásának első fázisában (4.12. ábra) a diákoknak fel kellett fedeznie a rendszert, azaz a három bemeneti változó értékeit szabadon változtatva és megfigyelve a kimeneti változók értékváltozását, fel kellett ismerni a probléma háttérében lévő összefüggésrendszert. A változók egymással való



kapcsolatát nyilak segítségével a probléma alatt található, a bementi és kimeneti változókat megjelenítő modellen meg is kellett jeleníteni. A válaszok kiértékelése során akkor kapott 1 pontot a diák, ha a rendszerben lévő összefüggések mindegyikét pontosan felrajzolta, azaz tökéletes modellt állított el. Ellenkező esetben 0 pontot kapott a probléma e fázisának megoldására. A tesztek egy részletét lásd a 8.3. mellékletben.

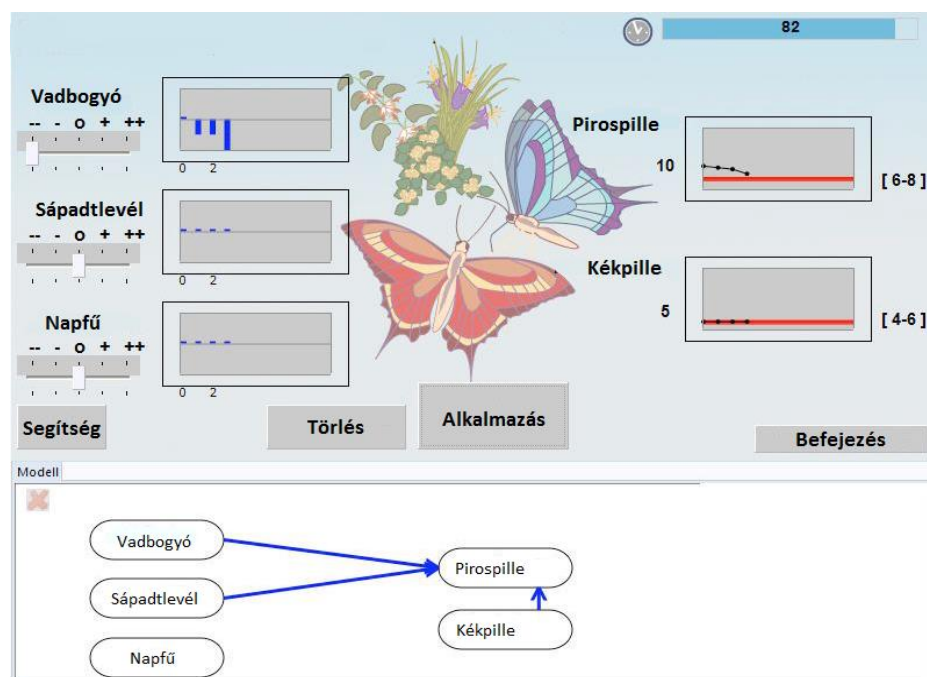


4.12. ábra

*Az eDia platformon a teszt egyik problémájának első része. (Egy hatalmas pillangóházban háromféle pillangót tenyésztasz: Pirospillét, Kékpillét és Zöldpillét. A pillangók sajnos nem úgy fejlődnek, ahogy te szeretnéd, ezért új virágokat telepítesz a pillangóházba: vadbogyót, sápadtlevelet és napfűvet. Ezek nektárja remélhetőleg segíti a pillangók fejlődését. Találd ki, hogy a vadbogyó, a sápadtlevél és a napfű nektárja milyen hatással van a különböző pillangófélék fejlődésére!)*

A problémák második fázisában működtetni kellett a rendszert (4.13. ábra), azaz megismerve a valódi összefüggéseket (a program megjeleníti a problémamegoldó számára a helyes összefüggésrendszert), a bemeneti változók értékeit állítva elérni a kimeneti változók előre meghatározott célértékeit (részletesen l. Molnár, 2013a). Mindezt a problémák megoldójának maximum 4 lépésben és kutatástól függően 90 vagy 180 másodperc alatt kell elérni. A válaszok értékelése során kizárólag abban az esetben kapott az adott diák 1 pontot a probléma ezen részének megoldására, ha megadott időn és lépésszámon belül sikeresen elérte az összes kitűzött célértéket, ellenkező esetben 0 ponttal értékeltük teljesítményét.

Az első adatfelvétel során évfolyamtól függetlenül minden egyes diák ugyanazt a nyolc dinamikus, területfüggetlen problémát tartalmazó tesztet kapta, amit a Heidelbergi Egyetem kutatói (Greiff és Funke, 2010; Wüstenberg, Greiff és Funke, 2012) dolgoztak ki, majd egy közös kutatás keretein belül (l. pl. Greiff, Wüstenberg, Molnár, Fischer, Funke és Csapó, 2013; Molnár, Greiff és Csapó, 2012; R. Tóth, Molnár, Wüstenberg, Greiff és Csapó, 2011) megtörtént a teszt hazai adaptációja. A problémák mindkét részének megoldására 90 másodperc állt a diákok rendelkezésére. Az adatfelvétel a TAO platformon keresztül zajlott.



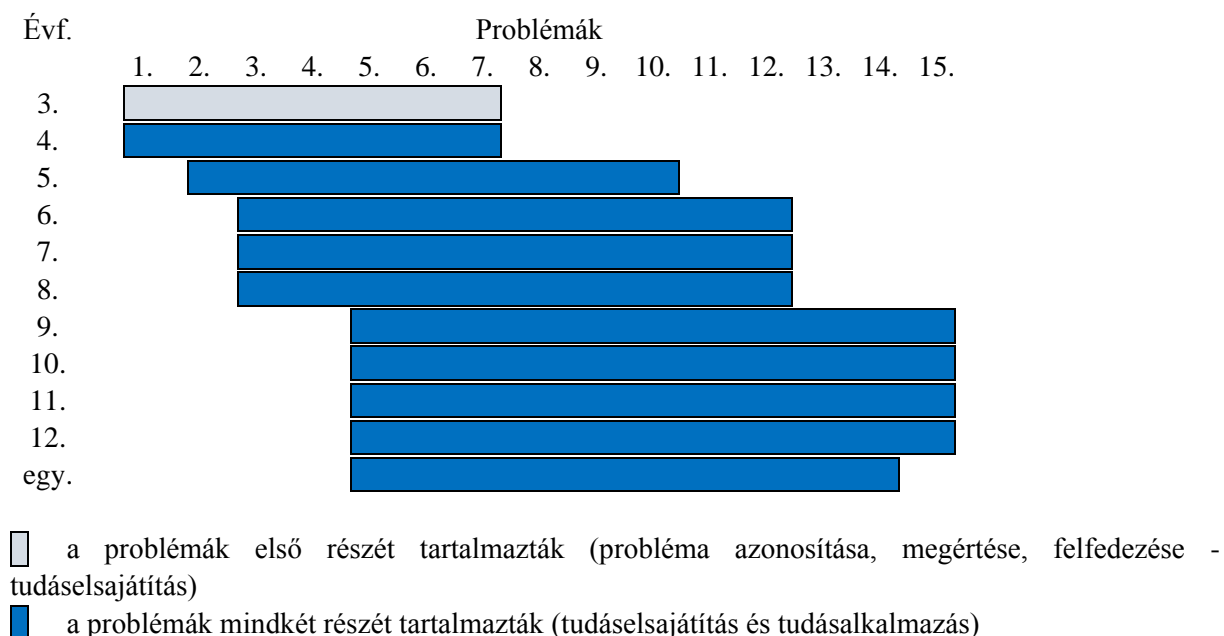
4.13. ábra

A TAO platformon a teszt egyik problémájának második része

A második, harmadik, negyedik és ötödik adatfelvétel már az SZTE OK fejlesztésű eDia rendszeren valósult meg, így lehetőségünk volt a számítógép-alapú tesztelés adta lehetőségek szélesebb körű kihasználására és az egyes kutatások során akár ugyanazon struktúrájú és kontextusú probléma más-más beállítások mellett történő kikövetítésére. Az eDia rendszer segítségével történt adatfelvételeknél a teszt egy magyarázattal egybekötött kezdő, úgynevezett mintafeladatot tartalmazott, amit évfolyamtól, korosztálytól függően 7-11 különböző nehézségű probléma követett. A második és negyedik adatfelvétel során kikövetített 15 különálló problémából kétféle horgonyzási technikát ötvözve (egyszerű horgony, kövér horgony, l. korábban, illetve Molnár, 2013b) öt különböző nehézségi szintű teszt került összeállításra. A tesztek horgonyzásának évfolyamonkénti bontásban ábrázolt szerkezetét mutatja a 4.14. ábra, az egyes adatfelvételeken áthidaló horgonyzási rendszert lásd a 2. mellékletben. A diákoknak 180-180 másodperc állt rendelkezésére egy-egy probléma első és második fázisának megoldására.

A második adatfelvétel során kikövetített problémáknál a modellépítés fázisában bármelyik változótól bármelyik változóhoz tudtak nyilat rajzolni a diákok, azaz akár a kimeneti változókat reprezentáló alakzatoktól a bemeneti változókat reprezentáló alakzatokhoz. A

sajátdinamika esetleges modellezése is lehetséges volt minden egyes probléma kapcsán. A problémákat a diákok minden esetben meg tudták hallgatni, kiküszöbölve az olvasási nehézségeket és az olvasási különbségekből eredő esetleges teljesítménykülönbségeket. Az adatfelvétel előtt felhívtuk az iskolák figyelmét arra, hogy a problémák megoldásához a diákoknak fejhallgatóra és internetkapcsolatra lesz szüksége.



4.14. ábra

#### A kutatás felépítése – a problémák horgonyzása

A harmadik, kizárólag hetedik évfolyamos diákok körében történt adatfelvétel elején a tanulók videó segítségével sajátíthatták el a rendszer és a teszt használatának módját, ugyanakkor a problémák szövegét nem hallgathatták meg. A problémák két fázisának megoldására 90-90 másodperc állt rendelkezésükre, míg az egész teszt megoldására egy tanítási óra. A diákok csak a bemeneti változókból tudtak nyilat, összefüggést indítani a kimeneti változók felé (kivétel sajátinamika esetén, ahol a kimeneti változóból indított nyíl önmagába mutat vissza), fordítva nem. Sajátinamika modellezésére is csak azon feladatok esetén volt lehetőség, amely előtt egy rövid ismertetőt hallgathattak meg a sajátinamika lehetséges előfordulásáról és felismerése esetén modellezése lehetőségéről.

A negyedik és ötödik adatfelvétel során a diákok próbafeladatok és részletes instrukciót adó videók segítségével sajátíthatták el a rendszer használatát, a problémák modellezése során a harmadik adatfelvételnél alkalmazott elrendezést alkalmaztuk. A tág életkori intervallumot átfogó kutatásban a második kutatás kutatási elrendezését, horgonyzási technikáit használtuk, valamint 180-180 másodperc állt a tesztet megoldók rendelkezésére a problémák felfedezésére és a rendszer működtetésére. Ebben az esetben is csak a bemeneti változókból, illetve sajátinamika esetén a kimeneti változóból tudtak nyilat, összefüggést indítani a kimeneti változók bármelyike felé. A negyedik adatfelvétel során meghallgathatóak voltak az adott instrukciók, a problémák szövegei (l. összefoglalóan 4.3. táblázat).

A második, harmadik, negyedik és ötödik kutatásban a diákok az interaktív problémamegoldó képességet mérő teszteken kívül kitöltöttek egy problémamegoldó stratégiákra fókuszáló kérdőívet, mely eredményeiről I. Molnár (2013c) tanulmányát.

#### 4.3. táblázat. Az öt adatfelvétel során alkalmazott szcenárió főbb jellemzői

Adatfelvétel éve	2011	2013 pilot	2014 longitud.	2014 nagy mintás	2015 egyetemista
Kiközvetítés platformja	TAO	eDia	eDia	eDia	eDia
Egy problémarész időkorlátja	90mp	180mp	90mp	180mp	180mp
Modellrajzolás	korlátozott	nincs korlát	korlátozott	korlátozott	korlátozott
Instrukció meghallgathatósága	nem	igen	nem	igen	igen
A rendszer működését bemutató videó	nincs	van	van	van	van

A 2011-es kutatásban a dinamikus problémamegoldó képességet vizsgáló teszt mellett felvételre került egy területspecifikus statikus problémamegoldó képességet vizsgáló teszt (I. korábban), egy intelligenciateszt, egy induktív gondolkodás teszt és egy háttérkérdőív is. Intelligenciatesztként a fluid intelligencia vizsgálatára alkalmas, a g-faktort leginkább jelző, kultúrafüggetlen CFT 20-R (*Culture Fair Test 20-R*; Cattell és Weiss, 1980) tesztet (Cronbach- $\alpha=0,88$ ), illetve a 3. és az 5. évfolyam ennek egy résztesztjét (Cronbach  $\alpha=0,86$ ) adaptáltuk és alkalmaztuk. A Cattell–Horn–Carroll (CHC) elmélet értelmében ugyanis a g-faktor a kognitív képességek egy közös általános faktora. A g-faktor a CHC-modell 3. szintjén helyezkedik el, ami közvetlenül tíz, a modell második szintjén lévő, átfogó kognitív képességet befolyásol. A CHC-modell első, legalsó szintjén a specifikus kognitív képességek helyezkednek el (McGrew, 2009). A CHC-elmélet nemcsak oktatási kontextusban, oktatással foglalkozó kutatók körében, hanem más, pszichológiai mérés-értékeléssel foglalkozó kutatásokban is relevánsnak bizonyult és jelentős figyelmet keltett. Miután több évfolyamot átfogó összehasonlítást végeztünk, a CFT teszt különböző évfolyamok közötti mérési invarianciájának tesztelésére is sor került. A folytonos változóknál alkalmazható *Maximum Likelihood*-közelítést (ML) használtuk, illetve a modellek illeszkedésének összevetését közvetlenül megtehettük a  $\chi^2$  értékek és a szabadsági fokok (df) különbségének segítségével.

A teszt szigorú értelemben is mérési invarianciát mutatott (az elemzés módszerének leírását I. korábban), nem volt szignifikáns különbség sem az egyre szigorúbb modellek közötti illeszkedésben, sem azok konfigurális modellilleszkedéshez való viszonyításában sem (4.4. táblázat). Ennek következtében a teszten mutatott teljesítmények évfolyamtól függetlenül akár nyerspont érték szerint is összehasonlíthatóak egymással, ugyanis minden egyes évfolyamon ugyanazon látens konstrukció mérésére került sor.

A diákok induktív gondolkodásának fejlettségi szintjét Csapó Benő (2001) széles körben alkalmazott induktív gondolkodás tesztjére alapozva mértük. A tág életkori intervallum miatt három, egymással horgonyzott tesztváltozatot alkalmaztunk. Az első és második szintű tesztek között 15, a második és harmadik szintű tesztek között 8 horgony item, továbbá 26 kövér horgony (mindhárom szinten előforduló item) szerepelt. A három változat összesen 72 itemet (első szinten 48, második szinten 49, harmadik szinten 50 itemet) tartalmazott. A

számanalógiák és számsorok részteszt feleletalkotó itemekből, míg a szóanalógiák feleletválasztó itemekből állt. (Részletesebb ismertetésük: *Csapó, 2001; Molnár, Greiff és Csapó, 2013*). A tesztek megbízhatósági mutatója mindhárom szinten magas volt (Cronbach- $\alpha=0,93, 0,94, 0,95$ ).

4.4. táblázat. A CFT teszt mérési invariancia-vizsgálatának illeszkedésmutatói

Modell	$\chi^2$	df	$\Delta\chi^2$	$\Delta df$	p	CFI	TLI	RMSEA
(1) Konfigurális invariancia	3,908	6				0,999	0,999	0,000
(2) Faktorsúlyok invarianciája	11,603	12	7,7	6	>0,10	0,999	0,999	0,000
(3) Metrikus invariancia	20,902	18	17,0	12	>0,10	0,997	0,997	0,032
(4) Szigorú metrikus invariancia	25,546	26	21,6	20	>0,10	0,999	0,999	0,000

A 2015-ös, egyetemisták körében történt adatfelvételben a problémamegoldó képességet mérő tesztrel együtt összesen hat tesztet oldottak meg a diákok, valamint kitöltöttek egy tanulási stratégiákra [a kérdések egy része megegyezett a PISA 2012-ben végzett felmérésének tanulási stratégiákat vizsgáló kérdőívének kérdéseivel (*Artelt, Baumert, Julius-McElvani és Peschar, 2003*)] és korábbi tanulmányi sikerességre vonatkozó háttérkérdőívet. A tesztek kidolgozását végző mérés-értékelési szakemberek a tesztek átlagos nehézségi szintjét a frissen érettségizett hallgatóktól elvárható tudásszinthez igazították. A szakos hallgatók differenciálhatósága érdekében tartalmaztak könnyebb és nehezebb feladatokat is. A magyar nyelv és irodalom teszt 8 résztesztet és 130 itemet, a matematika 5 témakör szerint összesen 66 itemet tartalmazott. A történelem teszt 162 iteme átfogta az ókor, középkor, újkor és jelenkor történelmét. A természettudomány 166 iteme között szerepeltek fizika, kémia és biológia feladatok. Az angol teszt B1 és B2-es szintű feladatai (80 item) olvasás és hallás utáni értést vizsgáltak. Az empirikus megbízhatósági mutatók (4.5. táblázat) alapján a kidolgozott tesztek megfelelnek a szakmai követelményeknek, a magas megbízhatósági értékek arra utalnak, hogy az elért eredmények és megállapítások általánosíthatóak. A technológiaalapú tesztelés körütekintő alkalmazásával ma már releváns igényként merül fel a teljeskörű mérés-értékelési rendszer, beleértve az Országos kompetenciamérést és az érettségi vizsgát is, objektívebb, gazdaságosabb, elektronikus alapra való átállítása.

4.5. táblázat. Az egyetemista bemeneti mérés során alkalmazott tesztek reliabilitásmutatói az itemkihagyásos elemzések után

Teszt	Itemek száma	Cronbach- $\alpha$
Magyar	126	0,90
Matematika	63	0,89
Történelem	161	0,93
Természettudomány	163	0,88
Angol	80	0,96

Az elemzések során a klasszikus tesztelméleti eljárások mellett a mérési és strukturális invariancia ellenőrzését, valamint strukturális egyenleteken alapuló útelemzéseket, úgynevezett SEM (Bollen, 1989) elemzésekkel végeztük. A konfirmatív célú elemzések (CFA) során ellenőriztük a hipotetikus kapcsolatrendszer meglétét, azaz azt, hogy az adataink mennyire támasztják alá a szakirodalom által felállított, előzetesen kialakított modellt, az adott modellnek mennyire felelnek meg az adatok. Az exploratív célú felhasználás során modellszelekciót hajtottunk végre, ahol különféle (egymáshoz hasonló) modelleket illesztettünk és kiválasztottuk az adatainkra legjobban illeszkedő modellt. A modellek illeszkedésének vizsgálatához  $\chi^2$  illeszkedésvizsgálatot, CFI, TLI és RMSEA mutatókat alkalmaztunk (leírásukat 1. korábban).

A tág életkori mintán alkalmazott, egymással horgonyzott feladatlapok eredményeit, azok közös skálán történő jellemzését, az adatok skálázását Rasch modellel végeztük. A diákok képességszintjének meghatározásához 'mle' (*Maximum Likelihood Estimate*) értékeket, míg az egy kohorszra vonatkozó átlagos képességszint meghatározásokhoz plauzibilis értékeket (pv) számoltunk. A logitskálán lévő értékeket a korábbi hazai (Molnár, 2012; Molnár, Greiff és Csapó, 2013) és nemzetközi kutatások (1. PISA-kutatás) fényében 8. évfolyamos (15 éves) diákok eredménye alapján lineáris transzformációval 500 pont átlagú és 100 pont szórású skálára transzformáltuk. A görbeillesztés során négyparaméteres logisztikus görbe függvényt használtunk.

#### **4.3. A dinamikus problémamegoldó képességet mérő harmadik generációs tesztek működése: reliabilitás, faktorstruktúra, mérési invariancia, a problémák skálázása**

##### **4.3.1. A dinamikus problémamegoldó képesség tesztek jóságmutatói**

A 15 különböző szerkezetű és kontextusú (problémánként 2 itemet tartalmazó), interaktív, dinamikus problémából összeállított harmadik generációs dinamikus problémamegoldó képesség tesztek (DPK) belső konzisztenciája minden egyes évfolyamon megfelelőnek bizonyult (l. 4.6. táblázat). (A disszertációban a dinamikus és interaktív kifejezéseket szinonimakén használom.)

Tesztenkénti (évfolyamonkénti) bontásban a harmadik évfolyamos diákok eredményei alapján számolt jóságmutató volt a legalacsonyabb (Cronbach- $\alpha$ =0,75) a pilot mérés során. Ennek oka a tesztben alkalmazott alacsonyabb itemszám, illetve az alacsonyabb teljesítmények. Részükre az eredetileg 15 éves diákoknak kidolgozott tesztfeladatok bonyolultsága miatt nem került kiközvetítésre a problémák második, tudásalkalmazást mérő része.

Összességében a pilot és nagymintás adatfelvétel során 4. évfolyamtól az egyetemista korosztályig megbízhatóan viselkedtek a tesztek, a reliabilitásmutató értéke 0,79 és 0,92 között mozgott. Ez alapján megállapítható, hogy már a negyedik évfolyamos diákok is felkészültek a harmadikgenerációs tesztekkel történő empirikus kutatásokra. Kidolgozhatóak olyan, a technológia adta lehetőségeket szélesebb körben kihasználó tesztek, melyek alkalmasak a diákok 21. században kulcsfontosságúnak számító problémamegoldó képességének interaktív környezetben való vizsgálatára.

## 4.6. táblázat. A DPK tesztváltozatok reliabilitásmutatói

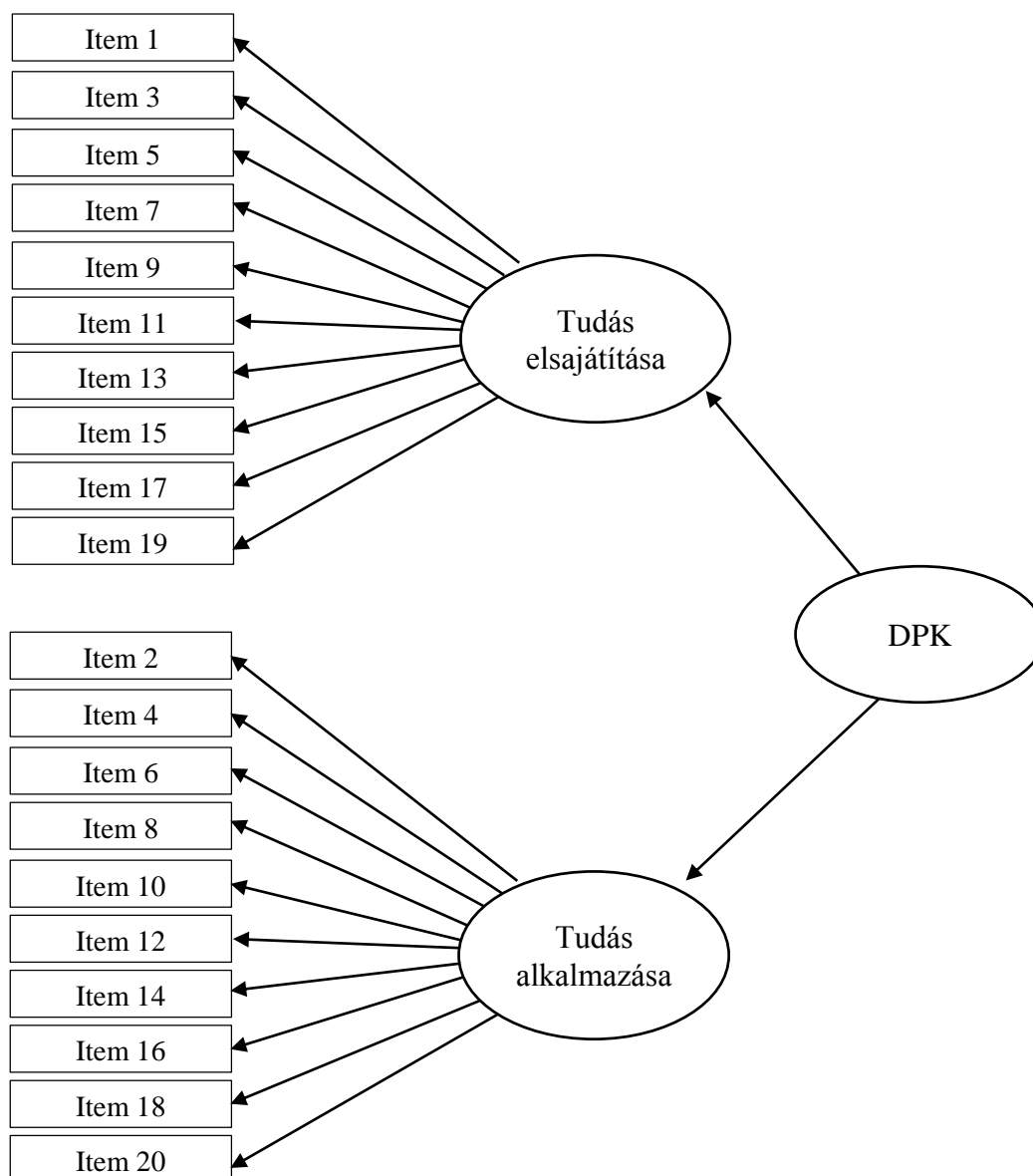
Adatfelvétel éve	Évfolyam	Itemszám	Cronbach- $\alpha$
2011	5-11.	14	0,92
2013	3.	8	0,75
	4.	16	0,82
	5.	20	0,85
	6-8.	22	0,87
	9.	24	0,87
2014	7.	18	0,81
2014	3.	6+1	0,88
	4.	14	0,79
	5.	18	0,78
	6.	20	0,80
	7.	20	0,83
	8.	20	0,82
	9.	22	0,90
	10.	22	0,90
	11.	22	0,90
	12.	22	0,90
2015	egyetemista	20	0,88

## 4.3.2. A problémamegoldó képesség teszt hipotetizált kétdimenziós modelljének empirikus verifikációja

A teszt problémáinak elméleti felépítése mögött egy kétdimenziós modell nyugszik (l. 4.2.2.2. részt), külön faktorként kezelve, egymástól elválasztva a tudás elsajátítás és a tudás alkalmazása faktorokat. Első lépésben konfirmációs faktoranalízist (CFA) alkalmaztunk a DPS modell ellenőrzésére, majd összehasonlítottuk a két beágyazott mérési modellt: az egyfaktoros (egydimenziós) és az elméleti modell által javasolt kétfaktoros (kétdimenziós) struktúrát, amely egy másodrendű faktorról rendelkezik (l. 4.15. ábra).

A modellben szereplő megfigyelhető, mérhető ún. manifeszt változókat konvencionálisan téglalappal, a nem mérhető ún. látens változókat ellipszissel jelöltük. A változók közötti kapcsolatokat nyilak szemléltetik. Az egyirányú nyilak ok-okozati viszonyt jelenítenek meg, míg a kétirányúak szimmetrikus, korrelációs viszonyra utalnak.

Az egyfaktoros modellt azért vettük bele az elemzésbe, mert a kétdimenziós modellen belül a két faktor látens szinten magasan korrelált egymással ( $r=0,74$ ;  $p<0,001$  – adatfelvételtől függetlenül, minden egye scenárióban 0,73-0,75 a két faktor közötti korreláció értéke; Greiff, Wüstenberg, Molnár, Fischer, Funke és Csapó, 2013). A modellek illeszkedésének teszteléséhez ismételtén több mutatót alkalmaztunk: az abszolút illeszkedési mutatót, a  $\chi^2$  értéket; a modell komplexitását figyelembe vevő, a megközelítés hibáját mérő RMSEA-t; illetve a relatív illeszkedési mutatókat a CFI és a TLI indexeket. Miután a teszt itemei dichotóm itemek, a modellillesztést WLSMV közelítési eljárással és THETA parametrizációval végeztük (Muthén és Muthén, 2010). A két modellilleszkedést pedig egy speciális  $\chi^2$ -próbával (DIFFTEST) hasonlítottunk össze.



4.15. ábra

*A dinamikus problémamegoldó képesség kétdimenziós (kétfaktoros) modellje (DPS: dinamikus problémamegoldó képesség)*

A tágabb életkori intervallum és az egyetemista mérésel azonos szcenárió miatt a 2014-es és 2015-ös adatfelvétel adatait összevonva használtuk ezen elemzésekben. Ezáltal egészen 4. évfolyamtól az egyetemista korosztályig vált lehetővé az alkalmazott dinamikus problémák viselkedésének vizsgálata. A tesztek horgonyzásának menetét követve alakítottuk ki az elemzésben alkalmazott életkori intervallumokat, ahol minden esetben összevetettük a két beágyazott mérési modell illeszkedését. Miután a SEM elemzések nagyobb mintaelemszámot igényelnek, mint a hagyományos manifeszt szintű statisztikai eljárások – *Ullman* (2007) javaslata értelmében a közelített paraméterek száma ne legyen több, mint a vizsgált minta ötöde (*Greiff, Wüstenberg, Molnár, Fischer, Funke és Csapó, 2013*) – ezért a 2014-es adatfelvétel 4. és 5. évfolyamos diákjainak adatait a közös itemekre alapozva, összevonva vettük bele ebbe az elemzésbe. (A 3. évfolyamosok tesztjében a problémák első fele jelent csak meg, így ott értelmezhetetlen volt a dimenzionalitásvizsgálat.)



Az illeszkedésmutatók minden esetben szignifikánsan magasabbak voltak a kétdimenziós modellnél (4.7. táblázat), a legmarkánsabb eltérés az egyetemista korosztályban realizálódott. Összességében beigazolódott az előzetes feltételezésünk és a hipotetizált kétdimenziós modell szignifikánsan jobban illeszkedett – a vizsgált életkortól függetlenül – az adatokhoz, mint az egydimenziós.

4.7. táblázat. A DPK teszt dimenzionalitás-vizsgálatának illeszkedésmutatói

Minta (évfolyam)	Modell	$\chi^2$	df	p	CFI	TLI	RMSEA
4-5	2 dimenziós	126,31	43	0,001	0,983	0,978	0,040
	1 dimenziós	150,19	44	0,001	0,978	0,972	0,044
6-8	2 dimenziós	536,74	169	0,001	0,973	0,970	0,031
	1 dimenziós	781,31	170	0,001	0,955	0,950	0,040
9-12	2 dimenziós	252,15	169	0,001	0,977	0,974	0,042
	1 dimenziós	256,71	170	0,001	0,975	0,973	0,043
Egyetem	2 dimenziós	410,81	169	0,001	0,971	0,967	0,044
	1 dimenziós	1072,05	170	0,001	0,906	0,891	0,080
Teljes minta	2 dimenziós	461,64	76	0,001	0,989	0,987	0,037
	1 dimenziós	1105,11	77	0,001	0,971	0,966	0,050

Az összevont elemzésbe a horgony itemek számának jelentős csökkenése miatt a legalacsonyabb korosztály adatait nem vettük bele, így a teljes mintára utaló adatok a 6. évfolyamtól az egyetemista korosztályig érvényesek ( $\chi^2$ -különbség teszt:  $\chi^2=235,085$ ;  $df=1$ ;  $p<0,001$ ). Az elméleti két (tudás elsajátítás és tudás alkalmazás) faktorra építő modell empirikusan megerősítésre került.

#### 4.3.3. A problémamegoldó képesség teszt mérési invarianciája

Az adatfelvételek tág életkori intervallumban, horgonyzott tesztekkel valósultak meg, ami lehetőséget biztosított az évfolyamonkénti teljesítmények összehasonlítására. A különböző életkorú diákok dinamikus problémamegoldó (DP) képességfejlettségi szintje összehasonlításának feltétele annak empirikus bizonyítása, hogy a teszt minden egyes évfolyamon, minden egyes vizsgált életkori intervallumban ugyanúgy viselkedett. Életkortól függetlenül ugyanazon konstruktum mérésére került sor, azaz az eredményeket nem befolyásolta más évfolyamspecifikus tulajdonság, kizárólag a képességfejlettségben lévő esetleges különbség.

A különböző scenáriók miatt minden olyan adatfelvétel esetén, ahol tág életkori intervallumban vontuk be diákokat a mintába, lefuttattuk az invarianciaelemzéseket. A 4.8. táblázat elemzései ezért magába foglalják a 2011, a 2013 és 2014-es adatfelvétel adatbázisain alapuló eredményeket.

A DPK teszt mérési invarianciájának tesztelése során mind a 2011-es, mind a 2013-as és mind a 2014-es adatfelvétel mindhárom egymásba ágyazott, egyre szigorúbb feltételeknek megfelelő modelljei jó illeszkedésűek voltak (CFI, TLI>0,95; RMSEA<0,06; l. 4.8. táblázat; az elemzései eljárás részletes leírását korábban, illetve Greiff, Wüstenberg, Molnár, Fischer, Funke és Csapó, 2013). A konfigurális modellhez képest egyik esetben sem romlott

szignifikánsan az illeszkedés mértéke sem a metrikus, sem a szigorú metrikus modell esetén ( $\Delta\text{CFI}<0,01$  és a  $\chi^2$ -különbség teszt nem szignifikáns), azaz a DPK teszt mérési invarianciája csoportok között empirikusan bizonyítottan tekinthető – függetlenül a mérés során alkalmazott szcenáriótól.

4.8. táblázat. A MicroDYN megközelítés fényében felépített DPK teszt mérési invariancia-vizsgálatának illeszkedésmutatói

Év	Modell	$\chi^2$	df	$\Delta\chi^2$	$\Delta\text{df}$	p	CFI	TLI	RMSEA
2011	Konfigurális invariancia	161,04	104	-			0,975	0,975	0,051
	Metrikus invariancia	170,10	115	22,29	23	>0,10	0,976	0,982	0,047
	Szigorú metrikus invariancia	165,82	116	53,16	43	>0,10	0,978	0,983	0,045
2013	Konfigurális invariancia	77,13	68				0,997	0,996	0,017
	Metrikus invariancia	87,78	74	10,52	6	>0,05	0,995	0,994	0,020
	Szigorú metrikus invariancia	105,08	84	16,34	10	>0,10	0,992	0,992	0,024
2014	Konfigurális invariancia	119,71	42	-	-	-	0,980	0,987	0,039
	Metrikus invariancia	126,33	45	7,37	3	>0,05	0,986	0,980	0,038
	Szigorú metrikus invariancia	145,49	52	15,02	8	>0,05	0,980	0,976	0,042

Az évfolyamonkénti átlagos különbségek mindhárom esetben valódi különbségek, melyek oka a diákok DP képességszintjében lévő esetleges eltérés és nem más pszichometriai tényező (Byrne és Stewart, 2006). Alkalmazhatóak és elvégezhetőek mindazon klasszikus és valószínűségi tesztelméleti elemzések, amelyek a különböző életkorú diákok DPK teljesítményének összehasonlításán alapulnak.

A nemek szerinti összehasonlítást a 2014-es, nagymintás vizsgálat adataira alapoztuk. A végső kétfaktoros mérési modellben a modifikációs indexek alapján az első faktorban két (d06-d08), a második faktorban két pár (d10-d11, illetve d11-d12) azonos faktorhoz tartozó item kapcsán engedték a korrelációt. A végső modell jól illeszkedett az adatokhoz (CFI=0,984, TLI=0,981, RMSEA=0,026,  $\chi^2=314,57$ , df=131). A három egymásba ágyazott, egyre szigorúbb feltételeknek megfelelő modellek illeszkedése sem romlott szignifikánsan (4.9. táblázat), azaz a fiúk és lányok teljesítménye összehasonlítható, mind a fiú, mind a lányok esetén azonos konstruktum mérésére került sor.

4.9. táblázat. A MicroDYN megközelítés fényében felépített DPK teszt nemek szerinti mérési invariancia-vizsgálatának illeszkedésmutatói

Modell	$\chi^2$	df	$\Delta\chi^2$	$\Delta df$	p	CFI	TLI	RMSEA
Konfigurális invariancia	416,94	264	-			0,986	0,984	0,024
Metrikus invariancia	428,73	278	21,32	14	>0,05	0,986	0,985	0,023
Szigorú metrikus invariancia	455,05	294	31,49	18	>0,05	0,985	0,984	0,023

#### 4.3.4. A problémák skálázása: a nehézségi index változása az alkalmazott scenárió függvényében

A 2013-as és 2014-es adatfelvétel eredményei alapján történt a problémák nehézségi indexének meghatározása. Miután a két adatfelvétel más-más scenárióban valósult meg, így ezen elemzések lehetőséget adtak arra, hogy feltérképezzük az azonos tartalmú, de különböző feltételek mellett alkalmazott problémák viselkedésének alakulását.

A 4.16. ábra bal oldali része a közös nehézségi, illetve képességszint skálát mutatja, a középső oszlopban elhelyezett számok a teszt egyes problémáit reprezentálják, a vertikális skálán elfoglalt helyük pedig a probléma nehézségére utal. Az ábra jobb oldali része az évfolyamonkénti bontásban megállapított átlagos képességszintet mutatja, minden egyes szám az azonos számú évfolyamot reprezentálja. A bal oldali oszlop a diákok képességszint szerinti eloszlását mutatja, ahol minden egyes 'X' nyolc diákot reprezentál.

Az évfolyamok és diákok átlagos képességszintjének függvényében végzett itemszintű elemzés eredménye alapján megállapítható, hogy a kidolgozott problémák 180-180 másodperc alatt történő megoldása, amikor bárholonnan bárhova húzhattak nyilat a diákok a modellépítés során (2. adatfelvétel kutatási elrendezése), összességében nehéznek bizonyultak. Annak ellenére, hogy az itemek nehézségi szintjei (-3,5 és +3,5 logitegység között) és az évfolyamok átlagos képességszintje (-1,5 és +1,0 logitegység között) azonos skálán mozogtak, arányaiban több olyan item szerepelt a tesztekben, amelyek alapvetően magasabb képességszint-tartományban (4.16. ábra) mértek igazán hatékonyan. Erre utal, hogy a diákok képességszint szerinti eloszlásgörbéje és a feladatok nehézségszint szerinti elhelyezkedése a közös skálán nem párhuzamos, inkább eltolt. Az évfolyamok átlagos képességszintjének elhelyezkedése viszont azt jelzi, hogy az 1 logitegységszint feletti tartományban is számos diák szerepel, csak egy konkrét képességszinten nincs belőlük nyolc, ezért nem jelenítette meg őket a program.

Ugyanezen elemzést lefuttattuk a 2014-es adatfelvétel adatain is (4.17. ábra), ahol továbbra is 180-180 másodperc állt a diákok rendelkezésére egy-egy probléma megoldásához, de ebben az esetben egyrészt a modellépítés során már nem húzhattak mindenholonnan mindenholra nyilat a diákok.

A sajátdinamikával nem rendelkező feladatok esetén a modellábrázolás során csak a bemeneti változókból indíthattak nyilat a diákok a kimeneti változók felé, fordított irányú nyíl rajzolását nem engedte a rendszer. A teszt elején három, a korábbiaknál is részletesebb videó mutatta be a rendszer használatát. Hasonlóan a 2013-as méréshez, egy példafeladat keretein

belül ki is próbálhatták a tesztelt személyek a feladatok egyes elemeinek működését. Az eredmények szerint a 2014-es adatfelvétel itemei e korlátozott scenárióban a korábbi év kutatásához képest erőteljesebben elkülönítették egymástól a diákokat, amivel párhuzamosan a feladatok is jobban szétszóródtak a nehézségi index skálán.

A két független skálázás eredménye közvetlenül, numerikusan nem összehasonlítható, miután a logitskálának nincs abszolút nulla pontja, illetve a logitegységnek abszolút hossza, de a feladatok egymáshoz való elhelyezkedése, viszonya már összevethető (l. korábban). Ezért a 4.16. és 4.17. ábra két személy/item térképét a feladatok egymáshoz való viszonyának vizualizálása érdekében összevetettük a 4.18. ábrán. Az összevetés során az azonos kontextusú feladatok azonos számmal szerepelnek, illetve az ábrán megjelenített feladatok közül kihagytuk azokat, amelyek nem szerepeltek mindkét adatfelvételben (ez némi eltérést eredményezett a 4.16., 4.17., illetve 4.18. ábrák között).

A három feladatból kettő modellrajzolósi probléma, azaz a problémák első részére vonatkozó item, amikor a diákoknak a rendszer megismerése után a rendszer viselkedésének modellezését kellett megvalósítani. Mindhárom itemben közös, hogy egyrészt 3 bemeneti és 3 kimeneti változó szerepelt a problémában, másrészt a 3-3 változó között négy reláció szerepelt, harmadrészt ezen itemek esetén sajátdinamika ábrázolása is lehetséges volt. A hipotetizált nehézségi viszonyokat, azaz, hogy a tesztrendszer legbonyolultabb feladatai tartoznak majd a legnehezebb feladatok közé jobban tükrözte a 2014-es scenárió, amikor az alapvetően öndinamikával nem rendelkező feladatok többsége esetén (amelyek a sajátdinamika rajzolását tanító mintafeladat előtt szerepeltek) nem volt engedélyezett a sajátdinamika rajzolása.

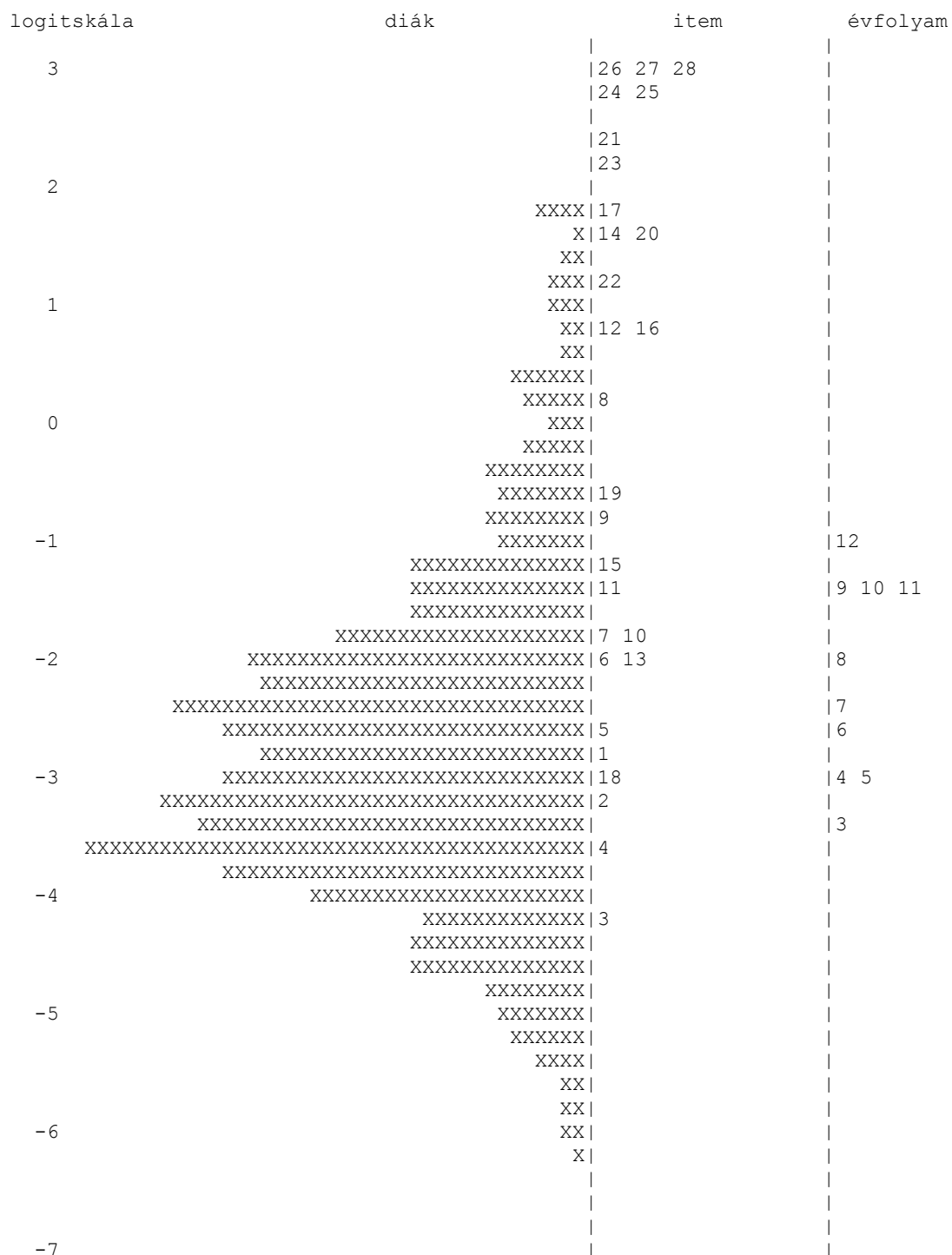
Összességében a 2013-as és 2014-es adatok alapján megállapítható, hogy a feladatokban történt korlátozás, igaz kis mértékben, de megváltoztatta a feladatok viselkedését. A változásban leginkább érintett feladatok közé tipikusan a legnehezebb és a legkönnyebb feladatok kerültek. Miután a 2014-es vizsgálatban korlátoztuk a helytelen válaszok körét (kimeneti változóból nem indulhatott nyíl bemeneti változó felé), ezért alapvetően több helyes válasz született, a feladatbank legkönnyebb feladatai is szélesebb nehézségi spektrumon helyezkedtek el. A legnehezebb feladatok viselkedésében történt változás hatására az előzetes feltevéseinknek megfelelő sorrendbe kerültek az itemek (a sajátdinamikával rendelkező feladatok átkerültek a legnehezebb feladatok közé). Ennek és a teszt megbízhatósága növekedésének oka nemcsak az itemek működésében történt változás, hanem az adatfelvétel nagyobb mintaelemszáma is lehet.

A továbbiakban a hetedik évfolyamos diákok válaszai alapján részletesen összehasonlítjuk a 2011-es (1), a 2013-as pilot (2), valamint a 2014-es longitudinális (3) és nagymintás (4) adatfelvétel közös feladatainak viselkedését. Mind a négy adatfelvételben más-más scenárió mellett kerültek kiközvetítésre a feladatok (részletesen l. 4.3. táblázat). Az elemzések korlátja egyrészt a kutatások mintájának különbözősége, miután mindegyik kutatás más-más mintán valósult meg, ezért ezen finomfelbontású eredményeket mintaspecifikus tulajdonságok is befolyásolhatták. Az elemzésbe csak azon diákok adatait vontuk be, akik mind a nyolc itemre választ adtak. Másrészt a négy adatfelvétel mindegyikében négy probléma, azaz nyolc közös item szerepelt, melyek közül három probléma a legnehezebb problémák közé tartozott. Az első probléma 2-2, a második, 3-2, a harmadik és negyedik pedig 3-3 bemeneti és kimeneti változót tartalmazott.

logitskála	item	évfolyam
	28	
3	30	
	27	
	25	
2	20	
	26	
	6 8 24	
	31 32	
1	X  22 29	9
	X	7 8
	X	
	XX	6
	X  2	
0	XXX  13 19	
	XX  17	
	XXX	5
	XXX  4 23	4
-1	XXXX  3 14	
	XXXXXX  18	
	XXXX  5 12	
	XXXXX  9 16	3
	XXXXXXX  7 10 11 21	
-2	XXXXXX	
	XXXXXXX	
	XXXXXXXXXX	
	XXXXXXXXX	
-3	XXXXXXXXXX  1	
	XXXXXXXXX	
	XXXXXXXXXX	
	XXXXXXXXXX  15	
	XXXXXXXXX	
-4	XXXXXX	
	XXXXXXXXX	
	XXXXX	
	XXXXX	
	XXXXX	
-5	XXX	
	XX	
	XX	
	X	
-6	X	
	X	
	X	
-7	X	

4.16. ábra

*Az itemek nehézségi indexei a diákok és az évfolyamok átlagos képességszintjének függvényében (a minta átlagos képességszintje 0-ra transzformált, minden egyes 'X' nyolc diákot reprezentál; részletesebben l. Molnár, 2013a; 2013-as adatfelvétel alapján)*



4.17. ábra

Az item-személy térkép az évfolyamok átlagos képességszintjének függvényében  
(minden egyes 'x' 8 diákot reprezentál)

*A 2013-as és 2014-es felmérés azonos tartalmú, de különböző feltételek mellett kiközvetített feladatai nehézségi szintjének alakulás egymás viszonylatában (minden egyes 'x' 8 diátok reprezentál)*

A nyolc itemre végzett megbízhatósági elemzés (4.10. táblázat) eredménye alapján a legelső adatfelvétel szcenáriója bizonyult a legjobbnak. Ugyanakkor figyelembe véve a diákok ezen itemeken nyújtott teljesítményét, valószínűsíthető, hogy a magasabb reliabilitásmutatók a minta magasabb, modellált átlagos képességszinthez közelebb álló képességszintjével is összefüggésben voltak. Ennek következtében több információt is szolgáltatott az adott item megoldása a diákok képességszintjéről, miután jobban, pontosabban differenciálta őket.

Az itemenkénti helyes megoldásokat adatfelvételi bontásban görcső alá véve megállapítható (4.19. ábra), hogy alapvetően az első adatfelvétel eloszlási mintázata különbözött a többitől. Ez az eltérés foka miatt mintaszpecifikus tulajdonságoknak köszönhető, ezért a továbbiakban a második, harmadik és negyedik adatfelvétel eredményeire fókuszálunk.

4.10. táblázat. Hetedik évfolyamon a négy adatfelvétel közös itemei alapján számolt alapstatisztikai mutatók

Adatfelvétel	N	Cronbach- $\alpha$	Átlag (max: 8)	Szórás	t (p)
1	157	0,86	3,29	2,79	{3,4}<{2}<{1}
2	223	0,77	2,09	2,02	
3	2237	0,78	1,60	1,81	
4	607	0,80	1,60	1,93	

Megj.: A {}-ben lévő számok a mérések (adatfelvétel) sorszáma.

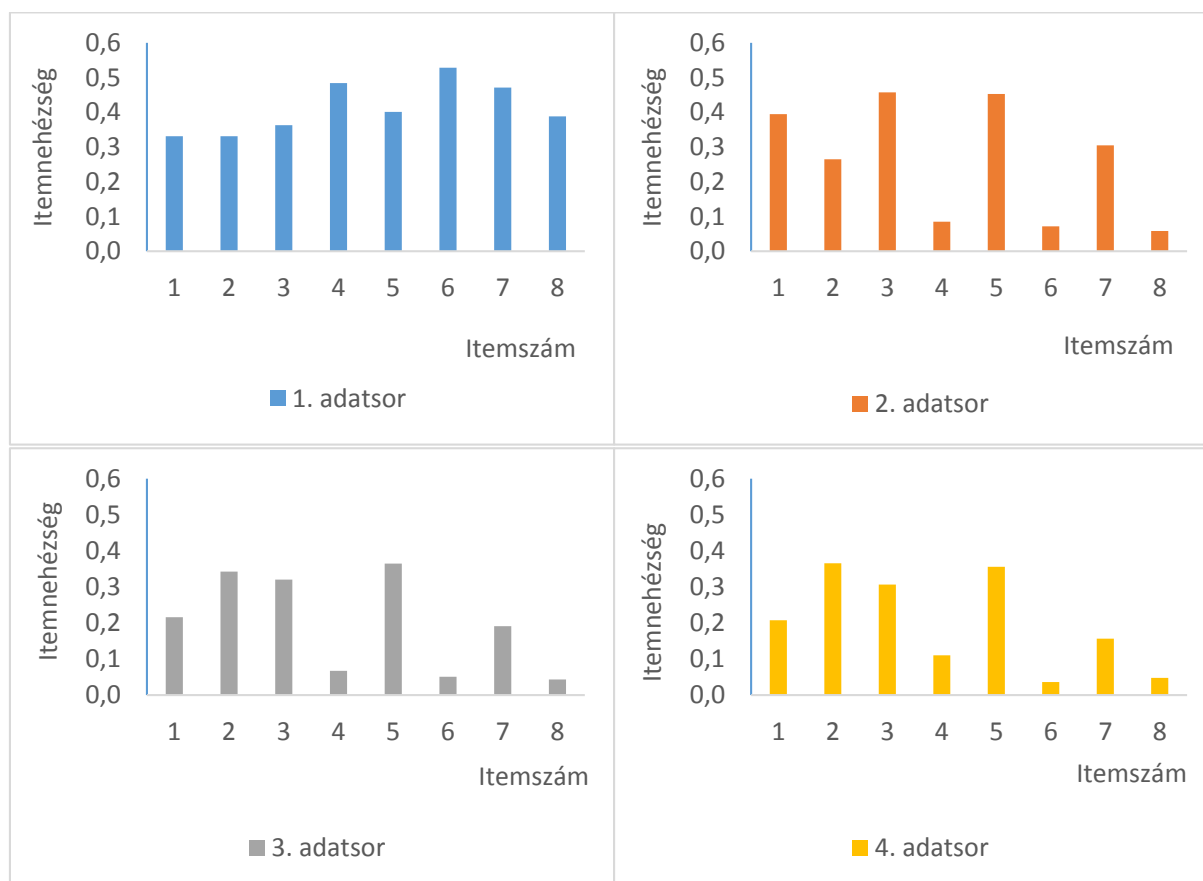
A második adatfelvétel során tapasztalt harmadik és negyedik adatfelvétel eredményeihez képest magasabb teljesítményt egyértelműen a páratlan sorszámú itemeken nyújtott teljesítmények okozták. A páratlan sorszámú itemek minden esetben a változók összefüggésének feltérképezésére és az összefüggések modellben történő ábrázolására vonatkoznak. A problémák második, azaz az alkalmazás részében már nem volt szignifikáns teljesítménybeli különbség kimutatható a második, harmadik és negyedik adatfelvétel adatai alapján. Utóbbi arra enged következtetni, hogy a problémák e részében történő két változtatás, a feladatok megoldására rendelkezésre álló idő – 90 vagy 180 másodperc –, illetve a feladatok meghallgathatósága nem bírt teljesítménybefolyásoló erővel e korosztály tekintetében.

A feladatok első részében nyújtott eltérő viselkedést (2. és 3., illetve 2. és 4. adatfelvétel eredményeit összevetve) már három környezeti feltétel megváltoztatása is okozhatta: időkorlát, meghallgathatóság és a modellrajzolósi feltételek megváltoztatása. Miután a harmadik és a negyedik adatfelvétel esetén nem volt különbség az itemszintű teljesítményekben, ezért a meghallgathatóság és az időkorlát nem okozhatta a teljesítménybeli különbségeket (a harmadik adatfelvételen 90 másodperc, a negyedikben 180 másodperc állt a diákok rendelkezésére, mindkettő korlátozott modellrajzolósi feltétellel). Hiába tölthettek több időt a diákok ezen itemek megoldásával, az nem vezetett magasabb teljesítményhez.

A logfájllelmzések megerősítik ezeket a hipotéziseket annak ellenére, hogy szignifikáns kapcsolat adódott a probléma egyes részein eltöltött idő és a probléma megoldásának sikeressége között is ( $r_{\text{modellépítés}}=0,187$ ,  $r_{\text{modellből}}=0,468$ ; mindkettő  $p<0,01$ ). Átlagosan a diákok kevesebb, mint 3%-a használta ki a rendelkezésére álló többletidőt a problémák megoldása során, ezt is a modellépítés (jelölés: modellépítés) és nem a célállapot elérését (jelölés: modellből) célzó fázisban tették (átlag<sub>idő\_modellepítés</sub>=51,8 sec<sub>idő\_modellepítés</sub>=18,3 sec; átlag<sub>idő\_modellből</sub>=28,2, sd<sub>idő\_modellből</sub>=12,38). Erős kapcsolat volt a probléma két részén eltöltött



idő mennyisége között, azaz átlagosan, aki a probléma modellépítés fázisában több időt töltött el, több időt szánt a megfelelő célállapot elérését célzó beállítások meghozására is ( $r=0,569$ ,  $p<0,01$ ). Elemezve a plusz időt kihasználó diákok teljesítményét, megállapítható, hogy a 90 másodperc feletti plusz idő felhasználása nem járt együtt hatékonyabb problémamegoldással. Ezt támasztják alá az erősödő korrelációs együtthatók is (ha az elemzésből töröljük a plusz időt használó diákok adatait; ( $r_{\text{modellépítés}}=0,225$ ,  $r_{\text{modellből}}=0,483$ ; mindkettő  $p<0,01$ ).



4.19. ábra

*A négy különböző szcenárióban megvalósuló adatfelvétel közös itemeinek itemnehézségi mutatója*

Ezért a modellrajzolást kívánó itemek (modellépítés fázis) eltérő működését a modellrajzolás korlátozása okozta (csak a bemeneti változókból lehetett nyilat rajzolni a kimeneti változók felé, fordítva nem). A feladatok viselkedésváltozását a második, harmadik és negyedik adatfelvétel vonatkozásában a négy itemre számolt reliabilitásmutatók értékének változásával jellemeztük (Cronbach- $\alpha=0,80$ ,  $0,75$ ,  $0,74$ ). Ezek alapján jobban működtek azok a feladatok, ahol nem alkalmaztunk korlátozást a modellrajzolás vonatkozásában.

A négy adatfelvétel során alkalmazott szcenárióváltoztatások közül a feladatok működését és képességszint-diagnosztizálási erejét nem befolyásolta az alkalmazott 90 másodperces, vagy 180 másodperces időkorlát, mindkettő azonos hatékonysággal működött. Hetedik évfolyamon nem befolyásolta az instrukciók meghallgathatósága sem a feladatok működését, feltételezhetően e feladattulajdonság főképp kisiskolás diákok körében bírt teljesítménybefolyásoló erővel (bár a meghallgathatóság mellett összességében minimalizáltuk

az esetleges olvasnivaló mennyiségét). Ezzel szemben a modellépítés során alkalmazott korlátozás már befolyásoló tényezővel bírt a feladat működése tekintetében, összességében jobban működtek azok a feladatok, ahol nem adtunk további segítséget a diákoknak azzal, hogy a kimeneti változókból nem lehetett kapcsolatot indítani a bemeneti változók felé.

A disszertáció negyedik fejezetében áttekintettük a problémamegoldó képesség mérésének elméleti hátterét, a kutatásokban alkalmazott problémák főbb ismertetőjegyeit (pl.: kontextus, ismerősség, komplexitás, jól definiáltság, dinamikuság, problémamegoldók száma szerinti megkülönböztetés). A szakirodalmi kitekintés alapján megállapítható, hogy mind hazai, mind nemzetközi szinten fokozatosan háttérbe kerültek a hagyományos adatfelvétellel is kivitelezhető, statikus problémákon alapuló kutatások és elterjedtek a számítógép-alapú adatfelvétel lehetőségeit kihasználó problémákat alkalmazó vizsgálatok. Szintetizáltuk a statikus környezetben történt, területspecifikus problémamegoldó képesség vizsgálatára irányuló hazai kutatásokat, melyek eredményeként felépítettünk egy 100 problémából álló, tág életkori intervallumban működő és használható itembankot, valamint jellemeztük a különböző típusú statikus problémákon nyújtott teljesítmények változását, a problémamegoldó képesség fejlődését. A negyedik fejezet második részében összegeztük a számítógép-alapú tesztelés lehetőségeit kihasználó dinamikus problémamegoldó képesség mérését célzó kutatásaink céljait, módszereit, valamint részletesen kitértünk a kutatásokban alkalmazott tesztek és problémák mérési hatékonyságára, működésére (reliabilitás, faktorstruktúra, mérési invariancia, a problémák skálázása szempontjából), különböző scenáriók mentén történő viselkedésének alakulására (pl. a problémák megoldásával töltött idő – *time on task* elemzések). Eredményeink szerint életkortól függetlenül megbízhatóan használhatóak a nemzetközi kooperációban kidolgozott tesztek. Az empirikus adaton nyugvó strukturális elemzések eredményei alátámasztották az elméleti modell felépítését. Az invarianciaelemzések eredményei megalapozták a különböző korú diákok teszten nyújtott teljesítményének összehasonlíthatóságát.

A disszertáció ötödik fejezetében mind fejlődési, mind strukturális perspektívából áttekintjük és összehasonlítjuk a tudás elsajátításában, alkalmazásában és transzferálásában kiemelt szerepet játszó gondolkodási képességek (problémamegoldó képesség és induktív gondolkodás) első és harmadik generációs tesztekkel történő mérési eredményeit, valamint számszerűsítjük az első generációs teszteken nyújtott problémamegoldó képességre vonatkozó teljesítmények előrejelző erejét a harmadik generációs problémamegoldó képességet mérő teszteredmények függvényében. Az ötödik fejezet második felében szintetizáljuk az interaktív helyzetekben (MicroDYN megközelítés adaptálásával) és tág életkori intervallumban mért problémamegoldó képesség fejlődésére, a fejlődést befolyásoló háttérváltozókra és a logfájl elemzésekre alapozó, a diákok tesztelés során mutatott problémamegoldó stratégiáit és azok változását feltérképező kutatási kérdésekre válaszoló elemzéseket és eredményeket. A negyedik és ötödik fejezetben ismertetett eredmények nemzetközi szinten is hiánypótlóak, ugyanis eddig főképp 15 éves korosztály tesztelésére alkalmazták a MicroDYN megközelítés fényében készített problémamegoldó képesség mérésre alkalmas feladatokat, alkalmazásukra kisiskolás diákok és felnőttek körében, illetve tág életkori intervallumban elvétve került sor. A bemutatásra kerülő, logfájl elemzéseken alapuló kutatási kérdések megválaszolására pedig akár néhány évvel ezelőttig a mérés-értékelés technikáinak fejlettségi szintje miatt nem is volt lehetőség.

## 5. A DINAMIKUS PROBLÉMAMEGOLDÓ KÉPESSÉG FEJLŐDÉSE ÉS A FEJLŐDÉST BEFOLYÁSOLÓ TÉNYEZŐK

Az 5. fejezetben bemutatott kutatási eredmények kivétel nélkül a 4. fejezetben ismertetett dinamikus problémamegoldó képesség mérésére fókuszáló vizsgálatok adatbázisain alapulnak. Az egyes elemzések kapcsán minden esetben feltüntetjük, hogy az öt adatfelvétel közül melyik mérés, vagy mérések adataira támaszkodunk. A 4. fejezet kapcsolatos elemzései a problémák és a tesztek működését, a technológiaalapú tesztek – beleértve a leginnovatívabb harmadikgenerációs tesztek is – 9-19 éves korban való alkalmazhatóságát járták körül. Az 5. fejezetben pedig a diákok problémamegoldó képességének fejlődése, a fejlődést befolyásoló és előrejelző tényezők feltérképezése, valamint a diákok problémamegoldó stratégiáinak feltérképezése áll a középpontban. Utóbbit a logfájelemzések alapján, a dinamikus problémamegoldó teszten mutatott viselkedés segítségével tesszük.

### 5.1. A dinamikus problémamegoldó képesség fejlődése

A dinamikus problémamegoldó képesség fejlettségi szintjét mérő tesztek invarianciavizsgálatának eredményeképpen egymással összehasonlíthatóak a különböző életkorban mutatott teljesítmények. A megjelenő különbségek a képességfejlettségben lévő eltérések és nem más pszichometriai tényezők okai (l. 4.3.3. részt).

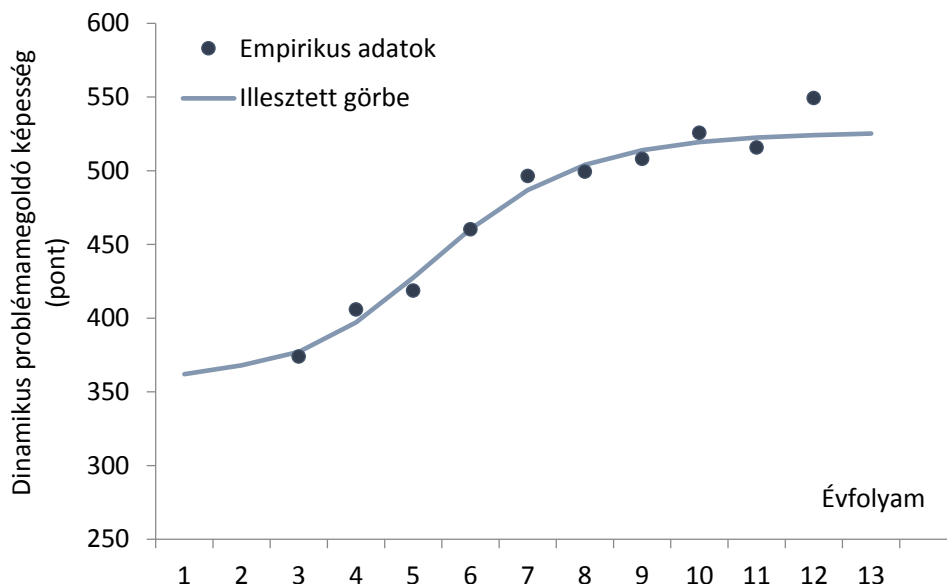
A dinamikus problémamegoldó képesség fejlődési görbéjének felrajzolását hasonlóan a korábbi, tág életkori intervallumot átfogó elemzésekhez, a 2014-es adatfelvétel adataira alapoztuk. Az eredmények megerősítik a 2013-as adatfelvétel alapján felrajzolt és prediktált fejlődési trendvonalat (l. Molnár és Pásztor-Kovács, 2015b).

A diákok évfolyamszintű átlagos képességszintjének meghatározása során minden egyes személyhez hozzárendeltünk egy plauzibilis értéket (*plausible values*, lehetséges értékek számolása, PV), mely a diszkrét pontokból álló képességszintszámolási eljárásokkal szemben (pl.: *weighted likelihood estimates*, súlyozott valószínűség, WLE; *maximum likelihood estimates*, legnagyobb valószínűség melletti közelítés, MLE; *expected a posteriori estimates*, várható értéken alapuló közelítés, a posteriori eloszlás átlaga, EAP) folytonos skálán helyezi el a diákok képességszintjének értékét (Molnár, 2013b). Ezt az eljárást populációsztintű paraméterszámolásra dolgozták ki.

A fejlődési folyamatok ismertetése során a 8. évfolyamos diákok eredményei alapján egy 500 átlagú és 100 szórású skálára transzformáltuk a fejlődést leíró görbét. A dinamikus problémák megoldottsága alapján megállapítható, hogy minden egyes évfolyamon nő vagy stagnál a diákok interaktív környezetben mutatott problémamegoldó képességének fejlettségi szintje (5.1. ábra). A fejlődés mértéke relatív lassú és évfolyamonként változó, átlagosan évenként a szórás ötödével fejlődik.

A legjelentősebb, az éves átlagos képességfejlődés kétszeresének megfelelő gyorsaságú fejlődés 5. és 7. (80 pont) évfolyam között van, míg 8. évfolyamon stagnálás tapasztalható. Harmadik évfolyamtól kilencedik évfolyamig összességében 134 pontos átlagos képességszintnövekedés figyelhető meg, ami csak kis mértékben haladja meg az

évfolyamonkénti bontásban számolt szórás átlagos (106 pont) értékét. Az empirikus adatokra illesztett négyparaméteres logisztikus görbe jól illeszkedik az adatokhoz ( $R^2=0,98$ ), ezért a görbe paraméterei alkalmasak e képesség fejlődésének jellemzésére. Az illesztett logisztikus görbe inflexiós pontja 6. évfolyamra tehető.



#### 5.1. ábra

*Az interaktív problémákon mutatott problémamegoldó képesség fejlődése 3-12. évfolyamon (2014-es adatfelvétel alapján)*

A dinamikus problémamegoldó képesség teszten nyújtott teljesítményeket összevetve a korábbi, papíralapú tesztekre alapozott problémamegoldó képességet statikus környezetben vizsgált eredményekkel:

- 1) négyparaméteres görbével jól modellezhető a képességfejlődés menete,
- 2) a fejlődés mértéke – hasonlóan más általános gondolkodási képesség fejlődéséhez (pl.: Adey, Csapó, Demetriou, Hautamaki és Shayer, 2007; Molnár és Csapó, 2011) – lassú,
- 3) 3–12. évfolyamok teljes időszakát tekintve évenként átlagosan a szórás ötödével fejlődik,
- 4) 4. évfolyamon minimális, 6. évfolyamon intenzív fejlődés figyelhető meg,
- 5) 6-7. évfolyamra esik a görbe inflexiós pontja (ebben az életkorban kiemelkedően hatékony a fejlesztés),
- 6) a 3. évfolyam előtt és a 12. évfolyam után is folytatódik e képesség fejlődése, több, azonos megállapításhoz jutunk. A görbék tulajdonságai megegyeznek meredekségük, elnyújtott S alakú, logisztikus görbére emlékeztető alakjuk tekintetében, inflexiós pontjukban azonban eltérés tapasztalható. Míg a statikus és területspecifikus problémák megoldása kapcsán a legerőteljesebb teljesítményváltozás 7. és 8. évfolyamon figyelhető meg, addig a dinamikus és területfüggetlen problémák esetén ez egy évvel korábbra tolódik és 6. és 7. évfolyamon a legintenzívebb a fejlődés mértéke. Ebből arra következtethetünk, hogy bár mindkét környezet a problémamegoldó képesség fejlettségi szintjének mérésére fókuszált, mégis a problémák kontextusa és dinamizmusában lévő eltérés teljesítménybefolyásoló erővel bír, bizonyos

oldalról más-más, de sok tekintetben azonos konstruktum mérését valósítják meg. Ennek részletesebb elemzésével, más gondolkodási képességeket is górcső alá véve a disszertáció későbbi részében foglalkozunk.

A statikus problémák megoldási sikeressége szerint arra a következtetésre jutottunk, hogy jelentős mértékben változnak a problémamegoldó képesség fejlődési trendvonalának egyes tulajdonságai (görbe meredeksége és maximuma), ha iskolatípusonkénti bontásban elemezzük az adatokat. Iskolatípustól függetlenül a fejlődés menete továbbra is jellemezhető volt egy négyparaméteres logisztikus görbével. Hasonló jelenség tapasztalható dinamikus problémák megoldása kapcsán is. Bár a szelekció nem a diákok problémamegoldó képességének fejlettségi szintje szerint történik, mégis jól jellemezhető a problémamegoldó képesség fejlettségi szintjével. Az általános iskola utáni szelekció jelentős mértékben növeli a különbséget a különböző iskolákba járó diákok között.

A középiskolai évfolyamokra vonatkozó elemzések korlátja, hogy relatív kis elemszámú diák vett részt évfolyamonként mind a 2013-as, mind a 2014-es adatfelvételben, ezért az iskolatípusonkénti bontásra alapozó megállapításainkat a 2011-es adatfelvétel eredményeire alapozzuk.

A 2011-es adatfelvétel szerint, amit alátámasztanak a 2013-as és 2014-es tapasztalatok is, a különbség mértéke a középiskolai évek alatt enyhén csökken, de nagysága az éves átlagos fejlődés mértékének többszöröse. A szelekció mértékét jellemzi, hogy egy 500 (100)-as skálán egyrészt 60 pont különbség van a 10. évfolyamos gimnazisták és a szintén érettségit adó szakközépiskolás kortársaik átlagos teljesítménye között. A gimnáziumban tanuló kortársaik átlagosan 600 pont körüli teljesítményt mutatnak. A szakiskolás 11. évfolyamos diákok átlagos teljesítménye a 8. évfolyamosok átlagos teljesítményszintjén van, több mint egy szórással alacsonyabban teljesítve gimnáziumba járó kortársaiknál. Extrapolálva a fejlődés folyamatát, megállapítható, hogy iskolatípustól függetlenül 11. évfolyam és a közoktatás vége után is, bár lassuló ütemben, de folytatódik a problémamegoldó képesség fejlődése.

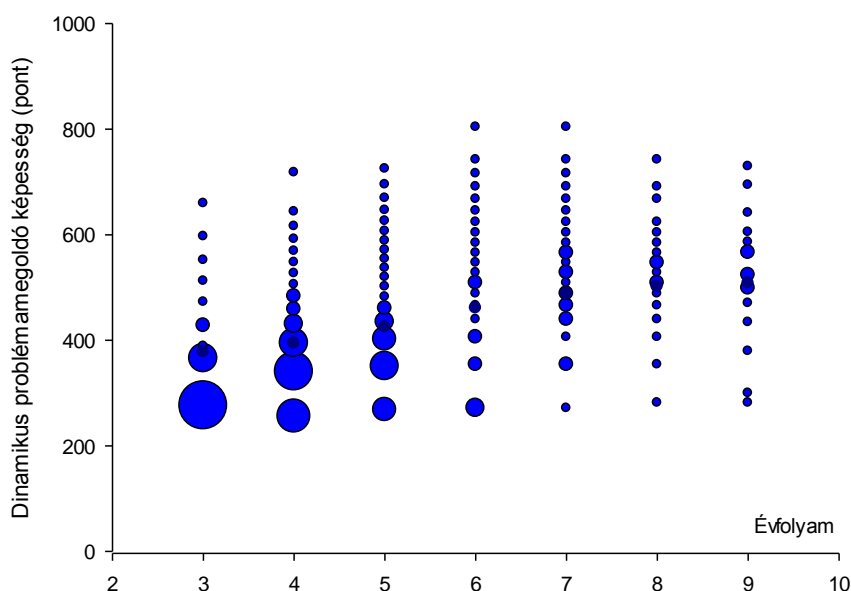
Adatfelvételi időponttól függetlenül a diákok között egy évfolyamon belül jelentkező képességszintbeli különbségek nagyobbak, mint a 3. és 12. évfolyamos diákok átlagos teljesítménye közötti különbség nagysága. Miután a középiskolás évfolyamok átlagosan 100 fő alatt képviseltették magukat a 2014-es adatfelvételben, ezért az ő adataikat alapvetően figyelmen kívül hagytuk az azonos évfolyamon belül jelentkező képességszintbeli különbségre vonatkozó elemzések során (5.2. ábra).

Minden egyes vizsgált évfolyamon a legalacsonyabb képességszintű diákok a 300 képességpont alatti tartományban vannak, ami alacsonyabb, mint egy átlagos 3. évfolyamos diák képességszintje (375 pont). Annak ellenére, hogy a legalacsonyabb képességszint-tartományban lévő diákok száma az évek előrehaladtával fokozatosan csökken, minden évfolyamon jelentkező lemaradók. A másik oldalról nézve, a legjobban teljesítő harmadik osztályos diákok (659 pont) szignifikánsan magasabb képességszintűek, mint egy átlagos nyolcadik (500 pont), vagy egy átlagos kilencedik (515) évfolyamos diák.

A 2014-es vizsgálat azonos feltételek, azonos scenárió alkalmazása mellett valósult meg, mint az egyetemisták körében történt 2015-ös adatfelvétel. A két mérés adatait összekészítve megállapítható, hogy, bár az egyetemre történő szelekció alapvetően nem a diákok problémamegoldó képességének fejlettségi szintje szerint történik, mégis a középiskolás átmenethez hasonlóan, ha rejtetten is, de jelentős szerepet játszik az egyetemre bejutásban. Az ország egyik vezető egyetemét kezdő, az adott évben érettségiző diákok képességszintje közös

képességskálán kifejezve a 2014-es vizsgálatban részt vevő 3-12. évfolyamos tanulókkal átlagosan 600,11 pont ( $sd=114,6$ ), azaz egy szórás távolságra van a 8. évfolyamos diákok átlagos képességszintjétől, jelentős mértékű fejlettségbeli különbségre utalva.

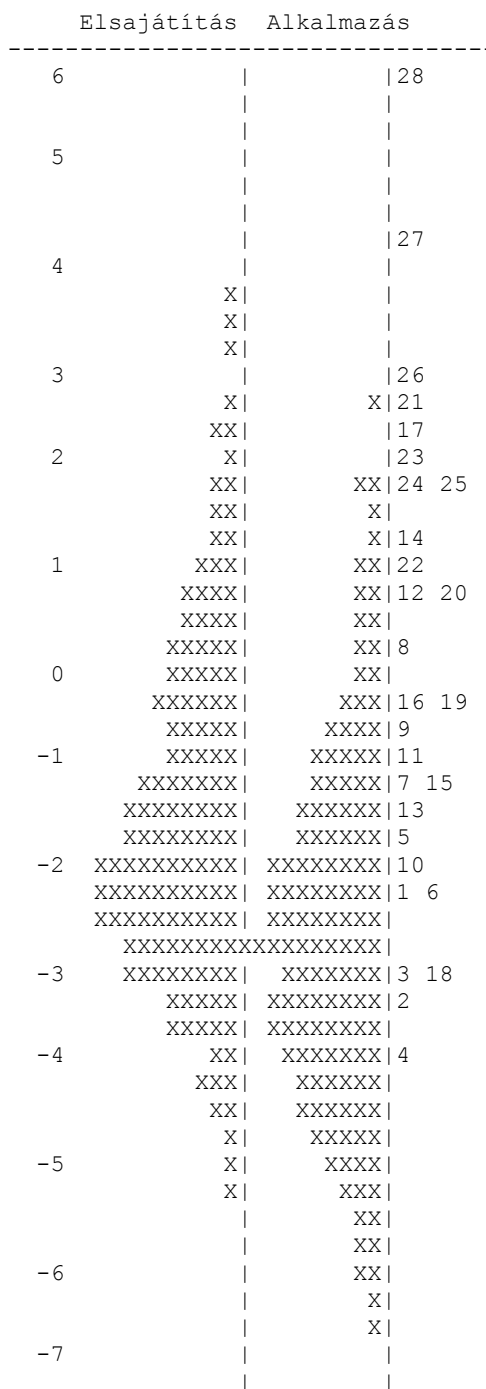
A dinamikus problémamegoldó képesség fejlődésére vonatkozó eddigi elemzésekben figyelmen kívül hagytuk, hogy a korábbi strukturális elemzések alapján a vizsgált konstruktum pontosabban jellemezhető egy kétfaktoros modellel. A tudás elsajátítása és a tudás alkalmazása faktorokhoz tartozó itemeken nyújtott teljesítményeket a teljes teszt szintjén eddig közösen kezeltük. Minden egyes diákhoz egyetlen egy képességszintet rendeltünk hozzá, ami közösen jellemezte a dinamikus problémák megoldásának két fázisában nyújtott összteljesítmény alapján problémamegoldó képességszintjüket. Ezt, a két faktor alapján külön-külön számolt képességszintek között lévő magas szintű korrelációs együttható miatt tehattük meg ( $r=0,74$ ,  $p<0,01$ ).



5.2. ábra

*A dinamikus problémamegoldó képesség fejlettségi szintjének eloszlása 3-9. évfolyamon*

A problémamegoldó képesség fejlettségi szintjére vonatkozó, a 2014-es és 2015-ös adatfelvétel eredményeit egyesítő kétdimenziós elemzés személy/item térképe alapján megállapítható, hogy bár mindkét dimenzióban, a dinamikus problémamegoldó képesség mindkét faktorában normál eloszlás tapasztalható (5.3. ábra), de a két görbe elhelyezkedése nem azonos szintű képességfejlettségre utal. A problémák megoldása során könnyebbnek bizonyultak a modellépítés, a tudás elsajátításának faktorába tartozó itemek megoldása, mint a tudás alkalmazása, egy adott célállapot elérése érdekében a szimulált, dinamikus rendszer működtetése ( $t=58,96$ ,  $p<0,01$ ).

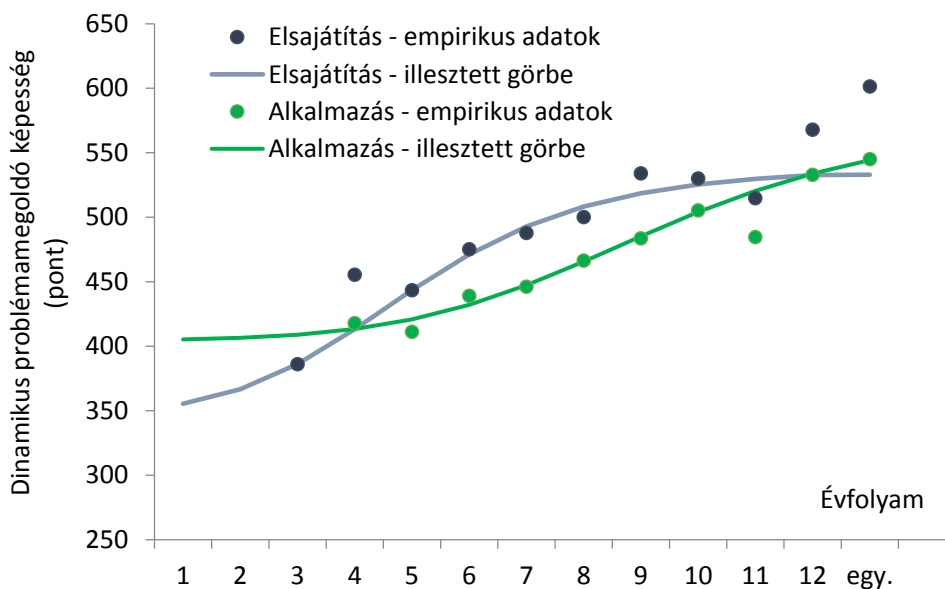


5.3. ábra

*A dinamikus problémamegoldó képesség teszt kétdimenziós személy/item térképe (A 2014-es és 2015-ös adatfelvétel egyesített adatbázisa alapján; minden egyes 'X' 35 diákot reprezentál.)*

A tudás elsajátítása és tudás alkalmazása terén tapasztalt fejlődési görbék eltérő tulajdonságokkal rendelkeznek (l. 5.4. ábra). A tudás elsajátítási szintjének fejlődését mutató empirikus adatokhoz történő görbeillesztés során figyelmen kívül hagytuk az egyetemista korosztály teljesítményét, miután eredményük jelentős mértékben eltért a 3-12. évfolyam teljesítményére illesztett görbe várható értékétől. Ennek következtében a tudás elsajátításának fejlődését leíró illesztett görbe illeszkedésmutatója  $R^2=0,97$ .

A tudás alkalmazását jellemző görbe illesztése során a 11. évfolyamos diákok adatai kerültek törlésre, viszont az egyetemista korosztály teljesítménye egyáltalán nem különbözött a közoktatásban tanuló diákok teljesítménye által felrajzolt és prediktált fejlődési szinttől. A görbe illeszkedésindexe  $R^2=0,99$ . (A 3. évfolyamosok tesztjében nem szerepeltek a tudás alkalmazásának faktorát mérő itemek.)



5.4. ábra

*A dinamikus problémamegoldó képesség két dimenziójának, faktorának (tudás elsajátítás és tudás alkalmazás) fejlődési görbéje (egy.=egyetem 1. évfolyam)*

A dinamikus problémamegoldó képesség két dimenziójának fejlődési görbéit összehasonlítva megállapítható, hogy egyrészt erős összefüggés van a fejlődési folyamatok között (amire a magas korrelációs együttható is utalt korábban). Minden, az adatfelvételbe bevont évfolyamon átlagosan magasabb képességszintűek a diákok a tudás elsajátítása, mint alkalmazása terén. A képességfejlettségben mutatkozó különbségek nagysága általános iskolában – 4-8. évfolyamon – kb. 40 pontnyi, ami a 4-8. évfolyamon tapasztható átlagos fejlődés nagyságának felel meg. Középiskolában 9. évfolyamon magasabb (50 pont), majd ez 11-12. évfolyamra 30-35 pontra csökken (a középiskolai eredmények általánosíthatóságának korlátja az alacsonyabb mintaelemszám, ami okozhatta a 12. évfolyamon tapasztalt, a korábbi adatokkal nem magyarázható képességszintnövekedést). Ez az eltérés szintén években mérhető fejlettségbeli különbségnek felel meg. Az egyetemisták körében megugrik a dinamikus problémamegoldó képesség két dimenziója között tapasztalt fejlődésbeli különbség nagysága (56 pont), ami a jelentősebb tudás elsajátítás faktorában mutatott teljesítménynövekedéssel magyarázható.

A két görbe inflexió pontja között is években mérhető a távolság. Míg a tudás elsajátítása kapcsán ez 5-6. évfolyamra, addig a tudás alkalmazása tekintetében 8-9. évfolyamra tehető. A fejlődés mértéke mindkét esetben nagyon lassú, utalva az explicit iskolai fejlesztés általános hiányára. Ez a fejlesztési hiány a tudás alkalmazása faktor kapcsán erőteljesebben jelentkezik.



A dinamikus problémamegoldó képesség tág életkori intervallumban történő fejlődésével kapcsolatos elemzések felhívták a figyelmet arra, hogy e képesség is, hasonlóan a többi gondolkodási képességhez (pl. induktív gondolkodás), az explicit iskolai fejlesztés hiánya miatt nagyon lassan fejlődik. Az egy évfolyamon belül tapasztalható fejlettségbeli különbségek nagysága meghaladja a 9-19 éves korosztályban történő fejlődés mértékét. Az iskolaátmeneti pontokon (általános – középiskola, középiskola – egyetem), ha implicit is, de egyértelmű a problémamegoldó képesség, mint tudás elsajátítás és alkalmazás mentén történő szelekció. A dinamikus problémamegoldó képesség mérése iskolai tananyagtól függetlenül hatékonyan megvalósítható számítógépes környezetben és tág életkori intervallumban (3. évfolyamtól az egyetemista korosztályig).

## 5.2. A problémamegoldó képességszint kapcsolata az intelligencia és az induktív gondolkodás fejlettségi szintjével, valamint az Országos kompetenciamérés eredményeivel, az iskolai sikerességgel és demográfiai háttérváltozókkal

### 5.2.1. A problémamegoldó képességszint kapcsolata az intelligencia és az induktív gondolkodás fejlettségi szintjével

Az itt ismertetett elemzések célja a különböző problémaszituációkban működtetett problémamegoldó képesség, az intelligencia és az induktív gondolkodás kapcsolatának feltérképezése fejlődés és együttjárás szempontjából. Az elemzések újszerűsége abban áll, hogy különválasztják a problémamegoldó képesség statikus és területfüggő, valamint dinamikus és területfüggetlen szituációkban történő működését, illetve tág életkori intervallumban elemzik az induktív gondolkodás, az intelligencia és a két különböző szituációban működtetett problémamegoldó képesség fejlettségi szintjének kapcsolatát.

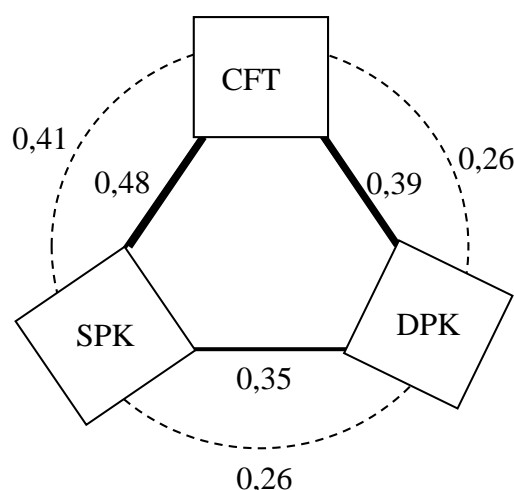
Az elemzések elvégzéséhez egy olyan kutatási elrendezésre volt szükség, ahol az adatfelvétel tág életkori intervallumban történik és minden egyes diák megold egy dinamikus és egy statikus problémamegoldás tesztet, valamint egy induktív gondolkodás és egy intelligenciatesztet is. Ez a típusú adatfelvétel diákonként legalább négy tanítási órányi mérést igényelt. A 2011-es adatfelvételünk felel meg ezeknek a kritériumoknak, ahol egy háttérkérdőív mellett a korábbi fejezetekben ismertetett statikus, valamint dinamikus problémamegoldás tesztek, illetve intelligenciatesztként a CFT 20-R figurális teszt, illetve *Csapó Benő* által kidolgozott induktív gondolkodás teszt került kiközzvetítésre. A tág életkori intervallum miatt a különböző évfolyamos diákok nem ugyanazt, de horgony itemekkel összekötött teszteket oldottak meg, melyek lehetővé tették az eredmények összehasonlítását, közös skálán történő jellemzését. A háttérkérdőív, az induktív gondolkodás (IG) teszt és a statikus problémamegoldás (SPK) teszt mind a kilenc évfolyamon (3–11.) kiközzvetítésre került. A dinamikus problémamegoldás (DPK) tesztet 5–11. évfolyamon alkalmaztuk, míg az intelligenciatesztet (CFT) 3., 5. és 7–11. évfolyamos diákok oldották meg. A kutatás mintájának részletes leírását, valamint az alkalmazott tesztek és működésük (megbízhatóság, mérési invariancia) elemzését lásd a 4.2. és 4.3. részben.

Az eredmények értelmében a diákok statikus problémák megoldása során nyújtott teljesítménye jelentős mértékben összefüggött intelligenciaszintjükkel ( $r=0,48$ ,  $p<0,01$ ). Az összefüggés szorossága erősebbnek bizonyult ( $z=2,11$ ,  $p<0,05$ ), mint az intelligencia és a dinamikus problémákon nyújtott teljesítményszint ( $r=0,39$ ,  $p<0,01$ ) kapcsolata, sőt erősebb

volt, mint a dinamikus és a statikus problémamegoldó teszten nyújtott ( $r=0,35$ ,  $p<0,01$ ) teljesítmények kapcsolata ( $z=3,84$ ,  $p<0,01$ ).

A parciális korreláció értékei (5.5. ábra) a lineáris (Pearson-féle) korrelációkénál minden esetben szignifikánsan alacsonyabbak voltak, azaz mindhárom kapcsolat erősségében meghatározó szerepet játszott a harmadik konstruktum fejlettségi szintje. Parciális korrelációk esetén is az intelligencia és a statikus problémákon nyújtott teljesítmények kapcsolata bizonyult a legerősebbnek, míg a másik két kapcsolat szorossága azonos volt.

Az átlagosan közepes erősségű korreláció jelenléte alátámasztja az utóbbi években végzett vonatkozó kutatási eredményeket, melyek közepes kapcsolatot bizonyítottak az intelligencia és a problémamegoldó képesség fejlettségi szintje között. Eredményeink értelmében az intelligencia jelentősebb szerepet játszik a területspecifikus és a statikus problémák megoldásakor, mint a területáltalános, dinamikus problémák megoldása közben. A dinamikus problémák megoldásának sikeressége azonos mértékben jelezhető előre az egyén intelligenciaszintjéből vagy a statikus problémákon nyújtott teljesítményéből.



5.5. ábra

*Az intelligencia és a problémamegoldó gondolkodás fejlettségének kapcsolata (minden koefficiens  $p<0,01$  szinten szignifikáns; CFT: intelligencia, SPK: problémamegoldó képesség – statikus problémák, DPK: problémamegoldó képesség – dinamikus problémák; a vonal mintázata a kapcsolat erősségét mutatja: szaggatott vonal: leggyengébb, vastag folytonos vonal legerősebb kapcsolat. A lineáris korreláció egyenes vonallal, a parciális korreláció görbe vonallal ábrázolt.)*

Az összefüggések erőssége változott évfolyamonkénti bontásban (5.1. táblázat). A statikus környezetben mutatott problémamegoldó képesség és az intelligencia vonatkozásában a kapcsolat szorossága az élekor előrehaladtával tendenciaszerűen nőtt, majd 7. évfolyam után állandósult. A statikus és a dinamikus problémahelyzetekben mutatott teljesítményekre a kezdeti függetlenség után egyre erősödő kapcsolat volt jellemző, majd szintén 7. évfolyam után az állandóság,  $r=0,40$  körüli korreláció, ami egybecsengett Molnár, Greiff és Csapó (2013) kutatási eredményeivel. A dinamikus problémamegoldás és az intelligencia kapcsolatát az 5. évfolyamon tapasztalt függetlenség után az állandó közepesen erős kapcsolat jellemezte. A  $z$ -

próba eredményei alapján 7–10. évfolyamon az összefüggés erőssége nem különbözött egymástól szignifikánsan.

Mindebből arra következtethetünk, hogy a statikus és a dinamikus problémahelyzetekben használt problémamegoldó módszerek, eszközök idővel egyre közelítenek egymáshoz, egyre hasonlóbbá válnak. Ha statikus problémamegoldó helyzetekben fejlesztjük a diákok problémamegoldó gondolkodását, stratégiáit, a fejlesztés mintegy melléktermékeként, transzferhatásaként a dinamikus problémamegoldó gondolkodásukat is fejlesztjük és fordítva. A transzferhatás mértékének pontosítása további kutatásokat igényel.

A diákok általános intelligenciájának az alsóbb évfolyamokon kisebb, majd egyre erősödő szerepe mutatkozott mind a területspecifikus és statikus, mind a területfüggetlen és dinamikus problémák megoldása kapcsán. Szerepe, közepes erősségű hatása mindkét típusú problémahelyzetben 7. évfolyam után változatlan. A fejlesztés szemszögéből az intelligencia és a problémamegoldó gondolkodás között lévő közepes erősségű korreláció arra utal, hogy ha fejlesztjük a diákok problémamegoldó gondolkodását, azzal általános intelligenciájuk is fejlődni fog.

5.1. táblázat. A statikus és a dinamikus problémahelyzetekben mutatott teljesítmény és az intelligencia kapcsolata évfolyamonkénti bontásban

Évfolyam	Statikus probléma és intelligencia	Statikus és dinamikus probléma	Dinamikus probléma és intelligencia
3	0,34**	–	–
4	–	–	–
5	0,25**	n.s.	n.s.
6	–	0,21**	–
7	0,49**	0,40**	0,33**
8	0,38**	0,36**	0,41**
9	0,41**	0,40**	0,54**
10	0,50**	0,23**	0,35**
11	–	0,43**	–

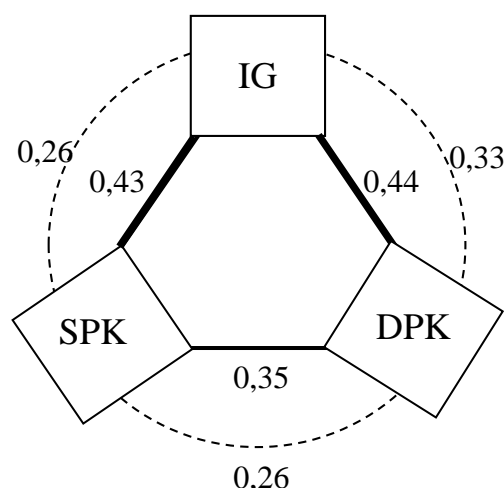
Megjegyzés: \*  $p < 0,05$ , \*\*  $p < 0,01$  szinten szignifikáns. A {} zárójelbe tett számok az évfolyamokat jelölik. Egy zárójelen belüli évfolyamok esetén nincs szignifikáns különbség a kapcsolat erősségében, míg a < jel a szignifikáns eltérésre, a kapcsolat erősödésére utal. –: a kutatási elrendezés miatt nincs adat.

Az induktív gondolkodás és a különböző problémahelyzetekben mutatott problémamegoldó képesség fejlettségi szintjének kapcsolata közepes volt (l. 5.6. ábra). A kapcsolat erőssége független volt az alkalmazott probléma dinamizmusától ( $r_{SPK}=0,43$  és  $r_{DPK}=0,44$ , mindkettő  $p < 0,01$ ), ugyanakkor az induktív gondolkodás fejlettségi szintje és a problémamegoldó képesség fejlettségi szintje közötti kapcsolat erősebbnek bizonyult ( $z=1,80$ ,  $p < 0,05$ ), mint a statikus és a dinamikus problémahelyzetekben nyújtott problémamegoldó sikeresség közötti kapcsolat ( $r=0,35$ ,  $p < 0,01$ ).

A parciális korreláció értékei a lineáris korrelációkénál minden esetben szignifikánsan alacsonyabbak voltak, azaz az intelligencia és a problémamegoldó képesség összefüggéseinek elemzése során is tapasztalt jelenséggel találkoztunk, mindhárom kapcsolat erősségében meghatározó szerepet játszott a harmadik konstruktum fejlettségi szintje ( $r_{IG\_SPK}=0,26$ ;

$r_{IG\_DPK}=0,33$ ;  $r_{SPK\_DPK}=0,26$ , mindegyik  $p<0,01$ ). A parciális korrelációs értékek azonos erősségű kapcsolatot jeleztek a három vizsgált terület között ( $p>0,05$ ).

Az eredmények értelmében az induktív gondolkodás, a statikus és a dinamikus problémahelyzetekben alkalmazott problémamegoldó képesség egymással összefüggő, de nem azonos konstruktumok. Az összefüggések erőssége összességében alátámasztotta a szakirodalomban megfogalmazódó elméleti kutatásokat (lásd pl.: Wirth és Klieme, 2004). Mindhárom vizsgált képesség olyan kognitív képességeket is működtet, amelyek nélkülözhetetlenek a szabályfelismerés, szabálygenerálás és szabályalkalmazás területén. Az induktív gondolkodás fejlettségi szintje, mint a problémamegoldás folyamatában az információfeldolgozás során működtetett alapvető gondolkodási képesség, erősebb hatással volt a különböző környezetben történő problémamegoldó sikerességre, mint a statikus problémahelyzetekben nyújtott problémamegoldó sikeresség a dinamikusra vagy fordítva. Ez alátámasztotta azon korábbi kutatási eredményeket (pl.: Klauer, 1996; Hamers, De Koning és Sijtsma, 2000), miszerint az induktív gondolkodás fejlettségi szintje alapvetően befolyásolja a tudás (idegennyelvtudást is; l. Csapó és Nikolov, 2009) elsajátításának és alkalmazásának sikerességét, ezért alapvető szerepet játszik mind a területspecifikus, mind a területfüggetlen, mind a statikus, mind a dinamikus problémák megoldása során.

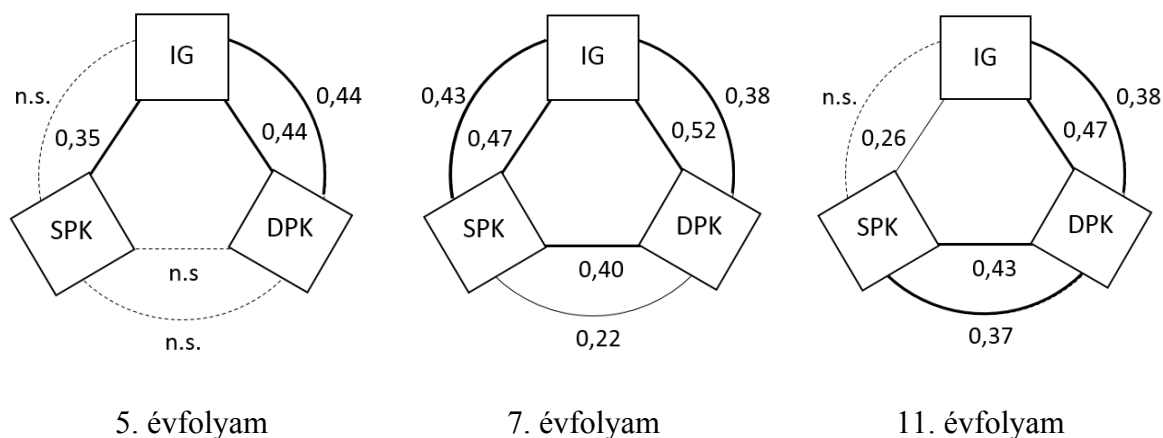


5.6. ábra

*Az induktív gondolkodás és a problémamegoldó képesség különböző típusú problémahelyzetekben mutatott fejlettségi szintje közötti kapcsolat (minden egyes együttható  $p<0,01$  szinten szignifikáns; SPK: statikus környezetben mutatott problémamegoldó képesség, DPK: dinamikus környezetben mutatott problémamegoldó képesség, IG: induktív gondolkodás; a kapcsolat erősségét jellemző vonalak magyarázatát l. 5.3. ábra)*

A kapcsolatok erősségének változásához a tesztek jószágmutatói, az illeszkedésindexek jósága és a képességfejlődés menete alapján három különböző életkort választottunk ki. Az 5. évfolyamos diákok fejlődés tekintetében még a gyors képességfejlődés fázisa előtt, a 7. évfolyamos diákok a fejlődés leggyorsabb fázisában, míg a 11. évfolyamos diákok átlagosan a gyors képességfejlődési periódus után vannak, közeledve a kötelező iskoláztatás végéhez. A korrelációs mintázat eltérő volt a három kiválasztott korcsoport esetén. Az összefüggések erőssége homogénabbnak bizonyult évfolyamon belül, mint évfolyamok között. A legerősebb

és évfolyamokat átívelően legstabilabb összefüggés az induktív gondolkodás és a dinamikus tesztkörnyezetben alkalmazott problémamegoldó képesség között volt. Az 5. és 11. évfolyamos diákoknál tapasztalt összefüggések erőssége egymáshoz hasonlóak, inkább alacsonyabbak voltak, míg a gyors képességfejlődés fázisában lévő 7. évfolyamos diákoknál tapasztaltuk a legerősebb lineáris és parciális korrelációs együtthatókat is (5.7. ábra).



5.7. ábra

*Az induktív gondolkodás és a különböző problémahelyzetekben alkalmazott problémamegoldó képesség fejlettségi szintje kapcsolatának változása 5., 7. és 11. évfolyamon (a vonalak vastagsága a kapcsolat erősségét mutatja, l. 5.3. ábra magyarázata; n.s.: nem szignifikáns kapcsolat)*

A különböző problémahelyzetekben működtetett problémamegoldó képességek sikeressége idővel egyre szorosabb kapcsolatban álltak egymással: alacsonyabb évfolyamon még nem volt szignifikáns kapcsolat a kettő között, míg a magasabb évfolyamokon már közepes, de szignifikáns volt a kapcsolat erőssége. A kilenc vizsgált kapcsolat esetén csak háromban voltak a parciális korrelációk értékei szignifikánsan alacsonyabbak, mint a lineáris korreláció értékei, ami arra utal, hogy nem feltétlen csökken a vizsgált konstruktumok közötti kapcsolat erőssége azzal, ha az összefüggések szorosságának megállapítása során kontrollálunk a harmadik konstruktumra.

Az induktív gondolkodás és a dinamikus problémahelyzetekben működtetett problémamegoldó képesség fejlettségi szintje közötti relatív stabil kapcsolat magyarázható az induktív gondolkodás más kognitív folyamatokban játszott szerepének elemzésével. Az induktív gondolkodás alapvető mechanizmusai, mint a tárgyak, azok tulajdonságainak és viszonyainak összehasonlítása, azonosságok és különbözőségek, valamint együttesen megjelenő azonosságok és különbözőségek felfedezése, továbbá ezek alapján a megjelenő szabályszerűségek felfedezése és alkalmazása, a dinamikus környezetben működtetett problémamegoldó képesség része is. Ahogy az 5.8. ábra korrelációs együtthatói jelzik, az induktív gondolkodás és a dinamikus helyzetekben történő problémamegoldó képességszint kapcsolata relatív magas és az is marad a statikus és területspecifikus helyzetben mutatott problémamegoldó sikerességre történő kontrollálás után is. Nem ez a helyzet az induktív gondolkodás és a statikus, területspecifikus helyzetekben működtetett problémamegoldó

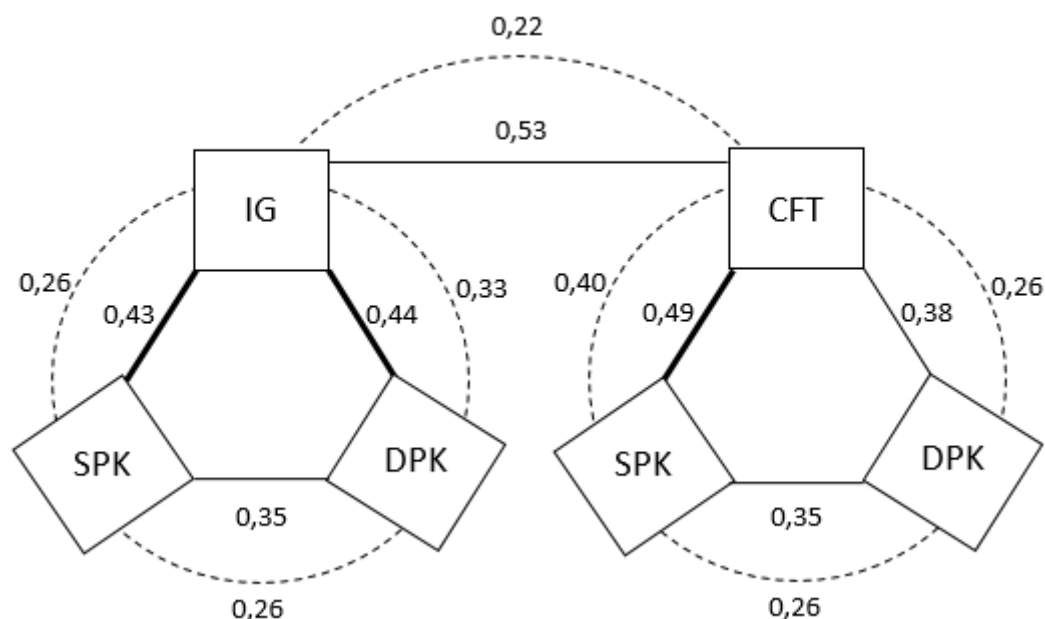
képesség fejlettségi szintje között, habár ez a jelenség az alacsonyabb és magasabb korcsoport esetén figyelhető meg, 7. évfolyamon nem tapasztható.

Az induktív gondolkodás területspecifikus és statikus helyzetekben mutatott problémamegoldó képesség fejlettségi szintjében játszott szerepe független az iskolában és iskolán kívül elsajátított információ mennyiségétől, csak a képességfejlődés menetének stádiumától függ. A hatás a képességfejlődés gyors fejlődési fázisában a legjelentősebb. Megfordítva, a kapcsolat erőssége azonos a gyors fejlődési fázis előtti és utáni stádiumban, függetlenül az iskolában elsajátított tudás mennyiségétől.

A különböző helyzetekben alkalmazott problémamegoldó képességek fejlettségi szintjének kapcsolata az időben egyre szorosabbá válik, ami azzal magyarázható, hogy a különböző problémahelyzetekben alkalmazott stratégiák egyre hasonlóbbá válnak. A területspecifikus, statikus helyzetekben történő problémamegoldás főképp a tudás alkalmazására épít, míg a területfüggetlen, interaktív problémakörnyezetben működtetett képességek előfeltételei az új tudás megszerzésének és alkalmazhatóságának, ami képességek működése szükséges területspecifikus statikus problémák megoldása során is.

Az induktív gondolkodás, területspecifikus, statikus és területáltalános dinamikus helyzetben működtetett problémamegoldó képesség fejlettségi szintjének kapcsolata hasonló volt, mint az intelligencia és a különböző problémahelyzetek megoldásának sikeressége közötti kapcsolat, a kapcsolatok erőssége  $r=0,35$  és  $0,49$  között mozgott (5.8. ábra). Az IG és az SPK vagy DPK ( $r=0,43$  és  $0,44$ ,  $p<0,01$ ), továbbá az intelligencia és a DPK ( $r=0,49$ ,  $p<0,01$ ) közötti kapcsolatok erőssége azonos volt, sőt erősebb ( $z=1,80$ ,  $p<0,05$ ), mint a DPK és az SPK ( $r=0,35$ ,  $p<0,01$ ) vagy az intelligencia és a DPK közötti kapcsolat erőssége ( $r=0,38$ ,  $p<0,01$ ). Az induktív gondolkodás és az intelligencia közötti kapcsolat bizonyult a legerősebbnek ( $r=0,53$ ,  $p<0,01$ ). Ennek oka, hogy az intelligenciatesztek legnagyobb része, így a jelen kutatásba alkalmazott CFT teszt is valójában induktív gondolkodást mér (pl: általánosítás, megkülönböztetés műveleteit). Az általunk alkalmazott induktív gondolkodás és intelligenciateszt közötti legnagyobb különbség abban rejlett, hogy előbbi iskolai kontextusban szóanalógiák, számanalógiák és számsorok alkalmazására, utóbbi figurális alakzatok tulajdonságai és a közöttük lévő kapcsolatok azonosságára, különbözőségére és együttesen megjelenő azonosságára és különbözőségére épített. Az induktív gondolkodás és az intelligencia közötti parciális korreláció értéke szignifikánsan alacsonyabb volt, azaz kapcsolatuk erősségét jelentős mértékben befolyásolta a diákok problémamegoldó képességének fejlettségi szintje.

Az eredmények összességében arra engednek következtetni, hogy ha a problémamegoldó képességet például területfüggetlen dinamikus környezetben fejlesztjük, akkor az statikus és területspecifikus környezetben is fejlődik, sőt bármely vizsgált konstruktum fejlesztése másodlagos hatásként pozitív transzferhatást gyakorol a többi vizsgált konstruktum fejlődésére. Egy esetleges gondolkodási képességfejlesztéssel pedig 7. évfolyamon érhetjük el a legnagyobb mértékű fejlesztést.



5.8. ábra

*Az induktív gondolkodás, az intelligencia és a különböző helyzetekben működtetett problémamegoldó képesség kapcsolata (minden együttható  $p < 0,01$  szinten szignifikáns; a vonalak és rövidítések magyarázatát l. 5.3. és 5.4. ábra)*

#### 5.2.2. A problémamegoldó képességszint kapcsolata az Országos kompetenciamérés eredményeivel, az iskolai sikerességgel és demográfiai háttérváltozókkal

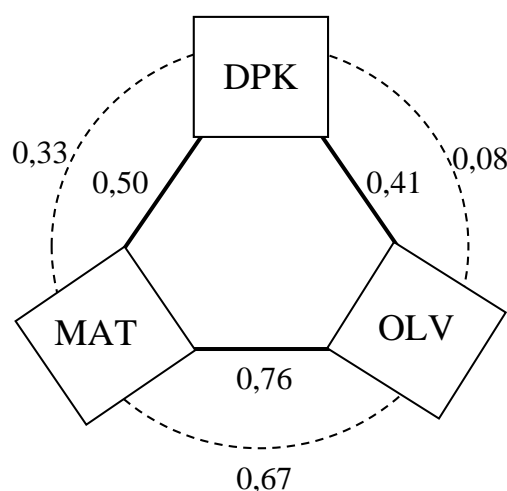
Az itt elvégzett elemzések adatbázisát a 2014-es hetedik, longitudinális adatfelvétel adta ( $n=2237$ ). Miután a diákok azonosítása mérési azonosítójuk segítségével történt, így lehetőségünk volt a 6. évfolyamos Országos kompetenciamérésen (OKM) elért eredmények és az OKM háttérkérdőívén adott válaszok lekérése után a DPK és az OKM adatbázisok egyesítésére.

A mérési invarianciaelemzés eredménye (l. korábban) alapján egymással összehasonlíthatóak a fiúk és a lányok DPK teszten nyújtott teljesítményei. Mind összteljesítményben, mind az egyes faktorok viszonylatában a fiúk átlagos képességszintje magasabb volt, mint a lányoké. A fiúk mind a tudás elsajátításának hatékonyságát, mind a tudás alkalmazásának sikerességét mérő itemeken jobban teljesítettek, mint a lányok (l. 5.2. táblázat).

Az Országos kompetenciamérés matematika és szövegértés tesztjén elért 6. évfolyamos eredmények közepesen-erős összefüggést mutattak a DPS teszten nyújtott teljesítményekkel ( $r_{DPS\_matek}=0,50$ ,  $r_{DPS\_szövegértés}=0,41$ ;  $p < 0,01$ ). A lineáris korrelációs értékek mindhárom esetben jelentős mértékben csökkentek, ha kontrolláltunk a harmadik változóra. A DPK és az olvasás teszten elért teljesítmény közötti gyenge kapcsolat a nagy mintaelemszám következtében maradt csak szignifikáns, de nem számottevő. Miután minden egyes összefüggés erőssége különböző (Fisher  $r$ -to- $z$  transzformáció alapján), az 5.9. ábrán eltértünk a korábban alkalmazott jelölési módtól és a lineáris korrelációt folytonos, a parciális korrelációt szaggatott vonalakkal jelöltük.

5.2. táblázat. A DPK teszten elért teljesítmények nemek szerinti bontásban

Terület	Nem	Átlag	Szórás	F	p	Welsch-féle d-próba	p
DPK	lány	4,11	2,87	53,69	<0,01	-6,05	<0,01
	fiú	4,93	3,54				
Tud. elsajátítás	lány	2,11	1,71	75,58	<0,01	-5,07	<0,01
	fiú	2,53	2,12				
Tud. alkalmazás	lány	1,99	1,58	28,41	<0,01	-5,71	<0,01
	fiú	2,40	1,81				



5.9. ábra

A DPK, az OKM matematika és szövegértés tesztjén elért teljesítmények kapcsolata (minden együttható  $p < 0,01$  szinten szignifikáns; lineáris korreláció: folytonos vonal, parciális korreláció: szaggatott vonat)

Az iskolai sikeresség és a DPK teszten nyújtott teljesítmény között gyenge-közepes kapcsolatot figyelhetünk meg ( $r=0,35$ ,  $p<0,01$ ). Tantárgyankénti bontásban a matematika és a történelem jegyek alakulása járt legszorosabban a DPK teszten elért teljesítményekkel ( $r=0,40$ ,  $0,33$ ;  $p<0,01$ ). Ezt követte a nyelvtan, a földrajz (mindkettő  $r=0,31$ ,  $p<0,01$ ), majd a szorgalom, az idegen nyelv és az irodalom (mindhárom  $r=0,29$ ,  $p<0,01$ ).

A diákok problémamegoldó képességszintje alacsony szinten függött össze édesanyjuk iskolai végzettségével ( $r=0,23$ ,  $p<0,01$ ), ami megerősíti a korábbi – területspecifikus és statikus problémahelyzetekre vonatkozó – kutatási eredményeket. A szülő iskolai végzettségénél meghatározóbb tényezőnek bizonyult a tervezett legmagasabb iskolai végzettség ( $r=0,28$ ,  $p<0,01$ ).

Összességében a vizsgált háttérváltozók mindegyike gyenge-közepes kapcsolatot mutatott a diákok problémamegoldó képességének fejlettségi szintjével, azaz iskolai jegyeik kevésbé jellemzik e fontos képességük fejlettségi szintjét. A diákok matematika teszten nyújtott teljesítménye, illetve iskolai matematika és történelem jegyük kapcsolata volt még a legerősebb. A DPK teszten elért teljesítmény alapján a fiúk – 7. évfolyamon – jobb



problémamegoldók, magasabb képességszinten vannak az ismeretek elsajátítása és alkalmazása terén, mint a lányok.

### **5.3. A gondolkodási képességek első generációs tesztekkel történő mérési eredményeinek előrejelző ereje a harmadik generációs teszttel mért problémamegoldó képességfejlettségi szintre**

A disszertáció előző részében a területfüggő statikus és területfüggetlen dinamikus helyzetekben működtetett problémamegoldó képességgel, valamint egyéb gondolkodási képességekkel kapcsolatos elemzések rávilágítottak arra, hogy közepes erősségű az összefüggés a két környezetben történő problémamegoldó sikeresség, valamint egyéb kulcsfontosságú, első generációs tesztekkel mért gondolkodási képességek fejlettségi szintje között. Mindezek alapján feltételezhetjük, hogy bizonyos mértékig lehetséges első generációs teszteredmények segítségével a harmadik generációs teszteken nyújtott teljesítmények előre jelzése, illetve fordítva is. Az első generációs tesztek egyirányú, oksági, előrejelző erejének számszerűsítéséhez strukturális egyenlet modellezés során végzett útelelemzéseket alkalmaztunk és teszteltük a feltételezett modelleket.

Az elemzések továbbra is a 2011-es adatfelvétel eredményein alapulnak, miután azok szükséges feltétele, hogy ugyanazon diákok kapcsán legyenek adataink első, illetve harmadik generációs teszttel mért problémamegoldó képességszintről, illetve az induktív gondolkodás és az intelligencia fejlettségi szintjéről, melyet szintén első generációs, hagyományos teszttel mértünk.

Az első, legegyszerűbb SEM modellben a diákok DPK teszteredményét kizárólag SPK teszten nyújtott teljesítményük alapján jeleztük előre. A mérési modell jól illeszkedett a vizsgált populációban ( $CFI=1,00$ ,  $TLI=1,00$ ,  $RMSEA=0,00$ ), a modell sztenderdizált útvonal-együtthatója 0,33 volt, a variancia jelentős része megmagyarázatlan maradt.

A második modellben a DPK teszten nyújtott teljesítmény előrejelzéséhez mind az SPK, mind a CFT (intelligencia) első generációs teszteken nyújtott teljesítményeket használtuk. A mérési modell továbbra is jó illeszkedésű volt ( $CFI=1,00$ ,  $TLI=1,00$ ,  $RMSEA=0,00$ ). A sztenderdizált útvonalegyütthatók értékei 0,19 (SPK/DPK) és 0,29 (CFT/DPK) voltak, de még mindig megmagyarázatlan maradt a variancia jelentős része.

A harmadik modellben mind az SPK, a CFT és az IG (induktív gondolkodás) első generációs teszteredményeket felhasználtuk a harmadik generációs DPK teszten nyújtott teljesítmény előrejelzése céljából. A sztenderdizált útvonal-együtthatók tovább csökkentek: 0,13 (SPK/DPK), 0,18 (CFT/DPK) és 0,26 (IG/DPK), jelezve az induktív gondolkodás kiemelt előrejelző szerepét. A modellilleszkedés továbbra is jó volt, de még ebben a modellben is megmagyarázatlan maradt a variancia jelentős része (*Molnár és mtsai, 2016*).

Az elemzések korlátja, hogy bár mind az SPK, mind a DPK tesztek a diákok problémamegoldó képességszintjét mérik, mégis szigorú értelemben különböző konstruktnak tekinthető (l. korábban) a két teszt által mért konstruktnak. Ezek alapján megállapítható, hogy bizonyos szintig első generációs teszteredmények alapján is előrejelezhető a diákok harmadik generációs teszteken nyújtott teljesítménye, holott a harmadik generációs tesztekkel történő munka több kognitív képesség működtetését igénylik, ezért a problémamegoldó képesség olyan aspektusait is vizsgálják, amik relevánsak a 21. század

gyorsan változó mindennapjaiban, de nem mérhetőek hagyományos, statikus, első generációs tesztek alkalmazásával.

### 5.3.1. Az induktív gondolkodás, az intelligencia és a szülők iskolai végzettségének előrejelző ereje a diákok problémamegoldó képességszintjére

Az induktív gondolkodás, az intelligencia és a szülők iskolai végzettségének a problémamegoldó képesség fejlettségi szintjére gyakorolt előrejelző hatását SEM útelemzésekkel végeztük. Az 5.2. részben kétirányú összefüggésekkel elemeztük a konstruktumok egymásra gyakorolt hatását, azonban egyirányú, oksági elemzésekre nem került sor. Itt a vizsgált konstruktumok mellett a nemzetközi vizsgálatokban jelentős háttérváltozónak számító szülő iskolai végzettségét is beemeltük az elemzésekbe.

A szakirodalom alapján felállított egyszerűbb elméleti modellben a szülők iskolai végzettsége a diákok induktív gondolkodásának fejlettségi szintjére bír előrejelző hatással (pl.: *Csapó, 2001*), ami, miután a tudás megszerzésének és alkalmazásának alapvető képessége, jelentős mértékben előrejelezheti a diákok problémamegoldó képességszintjét.

A mérési modell jó illeszkedésének bizonyult ( $CFI=0,89$ ,  $TLI=0,96$ ,  $RMSEA=0,03$ ). Az útvonal-együtthatók – a szülő iskolai végzettségének intelligenciára és statikus problémamegoldó képességre gyakorolt hatását kivéve – szignifikánsak voltak.

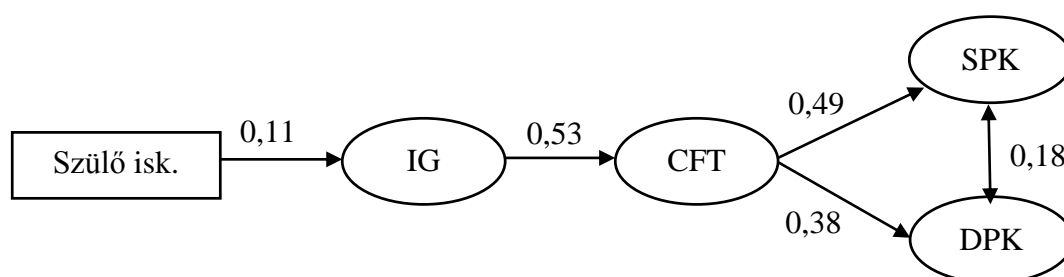
Az első modell útvonal-együtthatói alapján megállapítható, hogy a diákok induktív gondolkodásának fejlettségi szintje jelentős mértékben befolyásolja, előrejelzi problémamegoldó képességük fejlettségi szintjét, az induktív gondolkodásuk fejlettségi szintjét pedig kisebb mértékben, de prediktálja a szülők iskolai végzettsége. A sztenderdizált útvonal-együtthatók értéke: 0,19 (szülő iskolai végzettsége/IG), 0,43 (IG/SPK) és 0,41 (IG/DPK).

A második felállított és tesztelt modellben az induktív gondolkodás helyett az intelligencia jelenik meg, továbbá a szakirodalom alapján feltételeztük, hogy a szülők iskolai végzettsége prediktív erővel bír mind a diákok intelligenciaszintjére, mint problémamegoldó képességük fejlettségi szintjére. A mérési modell jó illeszkedésének bizonyult ( $CFI=0,97$ ,  $TLI=0,98$ ,  $RMSEA=0,02$ ). Az elemzések eredménye azonban nem mutatott szignifikáns oksági kapcsolatot a szülők iskolai végzettsége és a diákok intelligenciaszintje, valamint területfüggő, statikus környezetben mutatott problémamegoldó sikeressége között és enyhe oksági kapcsolatot jelzett a szülő iskolai végzettsége és a diákok dinamikus, területfüggetlen környezetben való problémamegoldó sikeressége között. Az útelemzés eredménye alapján a diákok intelligenciaszintje jelentős mértékben prediktálta problémamegoldó képességük fejlettségi szintjét. Kisebb mértékben a dinamikus, nagyobb mértékben a területfüggő, statikus környezetben történő problémamegoldó sikerességet.

Az első két modell eredményei alapján felépített harmadik, végső SEM modellben már helyet kapott mind a négy vizsgált konstruktum és az anya iskolai végzettsége is. Miután az intelligenciatesztek legnagyobb része, csakúgy, mint a jelen kutatásban alkalmazott CFT teszt induktív gondolkodást mérő feladatokat tartalmaz, ezért az induktív gondolkodás fejlettségi szintjét várjuk jelentős prediktív hatásúnak a diákok intelligenciateszten mutatott teljesítményére, ami várhatóan – a korábbi korrelációs értékek alapján – jól előrejelzi a problémamegoldó képesség fejlettségi szintjét.

A modell jó illeszkedésének bizonyult ( $CFI=0,99$ ,  $TLI=0,99$ ,  $RMSEA=0,01$ ), az útvonal-együtthatók viselkedése alátámasztják a szakirodalom és a korrelációs együtthatók alapján

megfogalmazottakat. A diákok intelligenciaszintje nagyobb mértékben jelzi előre az iskolai környezetben inkább előforduló, statikus, területfüggő kontextusban megvalósuló problémamegoldás sikerességét (sztdenderdizált útvonal-együttható: 0,49), mint a területfüggetlen, dinamikus környezetben történőt (sztdenderdizált útvonal-együttható: 0,38). A két problémahelyzetben történő problémamegoldás sikeressége továbbra is gyenge korrelációs együtthatót mutatott, jelezve az SPK és a DPK közös aspektusait, amik elkülöníthetőek az intelligenciától és a diákok induktív gondolkodásának fejlettségi szintjétől. Az induktív gondolkodás fejlettségi szintje pedig jelentős mértékű előrejelző erővel bírt a diákok intelligenciaszintjére (sztdenderdizált útvonal-együttható: 0,53), amit, bár kis mértékben, de előre jelzett a szülők iskolai végzettsége (5.10. ábra).



5.10. ábra

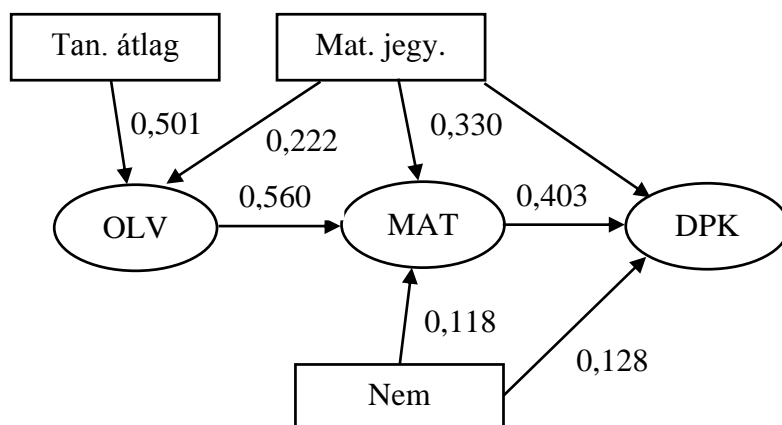
*A szülők iskolai végzettségének, a diákok induktív gondolkodása és intelligenciája fejlettségi szintjének előrejelző ereje a problémamegoldó képesség fejlettségi szintjére*

A korábbi korrelációs együtthatók és e modellek oksági útvonal-együtthatói alapján egyértelműen levonható az a következtetés, hogy az induktív gondolkodás fejlettségi szintje alapvető szerepet játszik a problémamegoldó folyamatokban. A különböző szcenáriókban működtetett problémamegoldó készségnek vannak közös elemei, működtetnek azonos képességeket, részképességeket, ugyanakkor az SPK és a DPK nem nevezhető azonos konstruktnak. Mindkét konstruktnak jól elkülöníthető az intelligenciától és az induktív gondolkodástól is. Utóbbi kiemelkedő szerepet játszik a diákok problémamegoldó képességének fejlettségi szintje tekintetében, azaz egy induktív gondolkodást fejlesztő tréning jelentős mértékben, transzferhatásként fejlesztené a diákok problémamegoldó képességét (szituációtól függetlenül), miután a problémamegoldó képesség egy olyan komplex képesség, aminek alapvető része az induktív gondolkodás, az induktív gondolkodás műveleteinek alkalmazása. Fejlesztéséről és fejlesztésének sikerességével, tartósságával kapcsolatos kutatásainkról l. Molnár (2011b, 2009b, 2008c, 2006c).

### 5.3.2. Az Országos kompetenciamérés, az iskolai sikeresség és a demográfiai háttérváltozók előrejelző ereje a diákok problémamegoldó képességszintjére

A disszertáció 5.2.2. részében ismertetett összefüggéselemzésekre építve a 2014-es longitudinális vizsgálat és a 2013-as OKM mérés adatbázisait egyesítve felépítettük az OKM által mért területek, az iskolai sikeresség és a demográfiai háttérváltozók előrejelző hatását magába foglaló modellt a diákok DPK teszten nyújtott teljesítményére. A végső SEM modell illeszkedése (CFI=0,99, TLI=0,99, RMSEA=0,04) jó volt.

A modellben a diákok matematika jegye, OKM matematika tesztjén nyújtott teljesítménye és neme által megmagyarázott variancia összességében alacsony volt ( $R^2=0,28$ ), azaz a DPK kapcsán a variancia jelentős része megmagyarázatlan maradt. Az előrejelző erő összességében mégis nagyobb volt, mint amit az intelligenciateszt és a statikus problémakörnyezetben nyújtott teljesítmény alapján számoltunk.



5.11. ábra

*Az Országos kompetenciamérés, az iskolai sikeresség és a demográfiai háttérváltozók előrejelző ereje a diákok problémamegoldó képességszintjére – útelemzés (minden útvonal-együttható  $p < 0,01$  szinten szignifikáns)*

Az OKM matematika tesztje a korábbiakban tapasztaltaknál erősebben jelzi előre a DPK fejlettségi szintjét. Prediktív ereje jelentősen nagyobb az iskolai jegyek alapján feltételezettekénél (a tanulmányi átlag nem bír szignifikáns előrejelző erővel a DPK fejlettségi szintjét tekintve ebben a modellben). Ennek oka, hogy az OKM matematika tesztjében előforduló feladatok döntő többsége gondolkodtató, problémamegoldást kívánó feladat volt és nem az iskolában tanultak „szó szerinti” visszaadását kérte. Erre utal a matematika jegy vártnál jóval alacsonyabb előrejelző ereje az OKM matematika tesztjére is. Az az eredmény, hogy a diákok tanulmányi átlagából egyáltalán nem következtethetünk DP képességük fejlettségi szintjére, mely képesség az OECD PISA szakértői által is kiemelkedő fontossággal bír a 21. században és amely képesség két azonosított faktora a tudás elsajátítás és tudás alkalmazás faktora, elgondolkodtató iskolarendszerünk hatékonyságát és működését illetően.

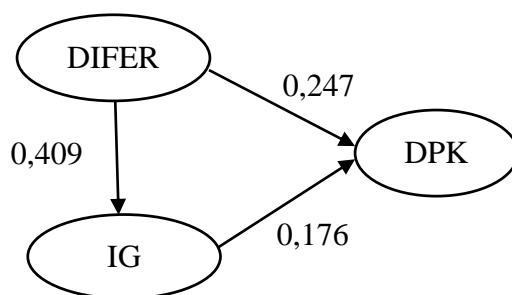
5.3.3. A diákok első évfolyamos DIFER, induktív gondolkodás, matematika és szövegértés teszteken nyújtott teljesítményének előrejelző hatása problémamegoldó képességük fejlettségi szintjére hetedik évfolyamon

A diákok első évfolyamos teljesítményeinek problémamegoldó képesség fejlettségi szintjére gyakorolt előrejelző hatását az MTA SZTE Képességfejlődés Kutatócsoport Longitudinális Program keretein belül felvett adatokkal végeztük. A 2014-ben 7. évfolyamos diákok (2014-es PDK adatfelvétel) 2007-ben kezdték az iskolát. A 2007/2008-as tanév kezdetén felvételre került a DIFER teszt, valamint az első év végi tesztelési periódusban egy szövegértés, egy számolási készség és egy induktív gondolkodás teszt. Az elemzésbe bevont

tesztek megoldására egy-egy teljes tanítási óra állt a diákok rendelkezésére. Az első évfolyamos tesztek papíralapon kerültek kiközvetítésre.

A modellek felépítése során a fokozatosság elvét követve első lépésként a DIFER, majd az első évfolyamos induktív gondolkodás teszten elért teljesítményük prediktív erejét számszerűsítettük külön-külön, majd egy közös modellben. Mindhárom modell modellilleszkedése ( $CFI=1,00$ ,  $TLI=1,00$  és  $RMSEA=0,00$ ) tökéletes volt. Az útelemzés eredménye alapján mindhárom modell útvonal-együtthatói szignifikáns erejűek voltak. A diákok DIFER teszten mutatott teljesítménye (sztenderdizált útvonal-együttható) 0,32 szinten, induktív gondolkodásuk fejlettségi szintje 0,28 mértékben prediktálta problémamegoldó képességük fejlettségi szintjét külön-külön elemezve azok hatását. A megmagyarázott variancia értéke a két modellben összességében még alacsony ( $R^2=0,10-0,15$ ) volt.

Közös modellben prediktálva a problémamegoldó képesség fejlettségi szintjét hét év távlatában kismértékben csökkentek az útvonal-együtthatók értékei: 0,25 (DIFER/DPK), 0,18 (IG/DPK), utalva arra, hogy a két terület nem teljesen független egymástól ( $r=0,41$ ,  $p<0,01$ ). Miután a diákok a DIFER tesztet az iskolába lépés után közvetlenül oldották meg, ezért a harmadik modellben (5.12. ábra) a DIFER, mint az induktív gondolkodás teszten nyújtott teljesítmény egy prediktív faktora is szerepelt.



5.12. ábra

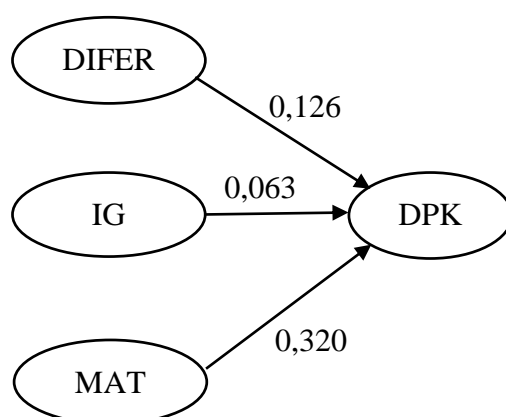
*Az első évfolyamos DIFER és induktív gondolkodás teszten nyújtott teljesítmények előrejelző ereje a hetedik évfolyamos problémamegoldó képességteszten nyújtott teljesítményre (minden útvonal-együttható  $p<0,01$  szinten szignifikáns)*

Az útvonal-együtthatók értékei alapján az előzetes hipotézisünknek megfelelően a DIFER jelentősebb mértékben jelezte előre az első évfolyamos induktív gondolkodás teszten nyújtott teljesítményt, mint a 7. évfolyamos dinamikus problémamegoldó képességteszten nyújtott teljesítményt, ugyanakkor a hetedik évfolyamos tesztre is gyakorolt prediktív hatást. A harmadik modellben továbbra is alacsony a megmagyarázott variancia értéke, az elsős (DIFER és IG) teszteredmények együttesen közel 13%-ban magyarázták a 7. évfolyamos problémamegoldó képességteszten mutatott teljesítményeket.

A negyedik modellben az iskolai tantárgyakhoz és iskolában tanítottakhoz közelebb álló két konstruktum, a szövegértés és számolási készség teszteken, első évfolyamon nyújtott teljesítmények prediktív erejét vizsgáltuk a hetedikes problémamegoldó képességteszten mutatott teljesítményekre. Előzetes hipotézisünknek megfelelően a szövegértés teszt eredménye nem bizonyult befolyásoló tényezőnek, miután az a hat évvel később felvett OKM tesztek segítségével végzett elemzésben sem volt szignifikáns. A számolás teszten nyújtott

teljesítmény viszont jelentősebb prediktív erővel bírt (sztenderdizált útvonal-együttható) 0,37, ami a DPK teszten nyújtott teljesítmény variációját 19%-ban magyarázta.

Az ötödik modellbe bevontuk mind a DIFER, mind az induktív gondolkodás, mind a számolási készség tesztek eredményeit. A diákok DIFER eredményével prediktáltuk induktív gondolkodás teszten, illetve matematika teszten nyújtott teljesítményét, majd mindhárom teszteredményt – a korábbi modellek értelmében – prediktív erőnek feltételeztük a problémamegoldó képesség teszten nyújtott hetedik évfolyamos teszteredményekre (5.13. ábra). A felépített modell illeszkedése azonban nem volt elfogadható, ezért a hatodik modellben mindhárom területet külön-külön kezelve építettük bele a modellbe, mely modell modellilleszkedése jó volt (CFI=1,00, TLI=1,00 és RMSEA=0,00), a modell által megmagyarázott variancia értéke  $R^2=0,20$ , azaz a korábbi modellekhez képest kismértékben növekedett. Az elemzésbe bevont változók útvonal-együtthatói szignifikánsak voltak.



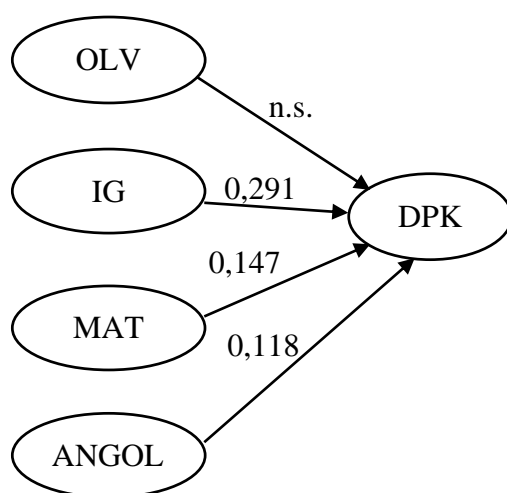
5.13. ábra

*Az első évfolyamos DIFER, induktív gondolkodás és számolási készség teszten nyújtott teljesítmények előrejelző ereje a hetedik évfolyamos problémamegoldó képességeszten nyújtott teljesítményre*

5.3.4. A diákok nemének, hatodik és hetedik évfolyamos induktív gondolkodás, angol, matematika és szövegértés teszteken nyújtott teljesítményének előrejelző hatása problémamegoldó képességük fejlettségi szintjére hetedik évfolyamon

A hét év távlatában végzett elemzések mellett ugyanezen diákok kizárólagosan hetedik évfolyamos eredményeire támaszkodva is elvégeztük ezeket az elemzéseket a célból, hogy elemezni tudjuk, milyen mértékben változik az egyes területek prediktív, meghatározó ereje a diákok problémamegoldó képessége fejlettségi szintje vonatkozásában az évek előrehaladtával. A felállított modellbe a diákok induktív gondolkodás, matematika, olvasás-szövegértés és angol teszteken mutatott, a longitudinális program keretein belül kidolgozott mérőeszközök segítségével történt adatfelvételek eredményeit vontuk be (5.14. ábra). A mérőeszközök felépítése mögötti keretrendszer eltért az OKM tesztekétől. Míg az OKM tesztek egyértelműen a tudás gondolkodási és alkalmazási dimenzióját mérik, addig a longitudinális vizsgálatba bevont matematika tesztben nagyobb arányban szerepeltek iskolai tudást mérő itemek. Az angol nyelv feltételezett prediktív hatását a szakirodalomra alapoztuk (Csapó és Nikolov, 2009).

A modellilleszkedés ( $CFI=1,00$ ,  $TLI=1,00$  és  $RMSEA=0,00$ ) jó volt. A diákok szövegértés teszten nyújtott teljesítménye – a vártaknak megfelelően – nem bírt prediktív erővel a problémamegoldó képességeszten nyújtott teljesítményükre, ami alátámasztja a korábbi OKM szövegértés teszt eredményei kapcsán tapasztaltakat. Ezúttal a matematika teszten nyújtott teljesítményük csak kis mértékben, míg az első generációs induktív gondolkodás teszten nyújtott teljesítményük nagyobb mértékben jelezte előre harmadik generációs problémamegoldó képességeszten nyújtott teljesítményüket, ami alátámasztotta korábbi vonatkozó eredményeinket. A matematika előrejelző fontossága az első évfolyamos teszteredményhez képest jelentős mértékben csökkent. Ennek oka lehet, hogy első évfolyamon még a diákok induktív gondolkodásának fejlettségi szintje nagyobb mértékben befolyásolta a matematika teszten elért teljesítményüket, mint hetedik évfolyamon. Az angol nyelvtudás foka, igaz kis mértékben, de hozzájárult a vizsgált konstuktumok által előrejelzett DP képességszint 26%-os megmagyarázott variáciájához.

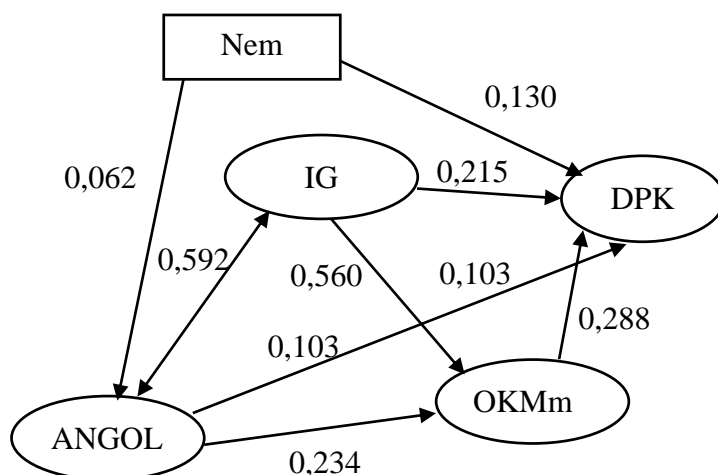


5.14. ábra

*A hetedik évfolyamos olvasás, induktív gondolkodás, matematika és angol teszteredmények előrejelző ereje a diákok problémamegoldó képességeszten nyújtott teljesítményre – útelemzés*

A disszertáció ezen részében bemutatott utolsó SEM modellben egyesítettük, egy közös modellben foglaltuk össze azokat a változókat, amelyek a korábbi modellek alapján előrejelző erővel bírtak a diákok DPK fejlettségi szintje kapcsán. A megfelelő illeszkedésmutatókkal rendelkező modellben ( $CFI=0,996$ ,  $TLI=0,977$ ,  $RMSEA=0,045$ ) a DIFER teszt, az iskolai jegyek (matematika és iskolai jegyek átlaga), az olvasás tesztek és a diszciplináris tudást vizsgáló matematika teszt eredménye nem mutatott szignifikáns előrejelző erőt a diákok DP képességének fejlettség szintjére (sem a modellbe bevont jelentős prediktív erővel bíró területek fejlettségi szintjére). Ezért az 5.15. ábrán szemléltetett modellben e változók nem kerültek megjelenítésre. Továbbra is az OKM matematika tesztje és a diákok induktív gondolkodásának fejlettségi szintje jelezte leginkább előre a DPK teszten nyújtott teljesítményt, melyek a nemmel és az angol idegen nyelvtudás szintjével együtt a DPK teszt variáciájának 30%-át magyarázták meg. A diákok induktív gondolkodásának fejlettségi szintje jelentős előrejelző hatást gyakorolt az OKM matematika tesztén elért teljesítményre, ami alátámasztja a teszt korábban említett

felépítését. A modell kiemeli az angol nyelvtudás és az induktív gondolkodás fejlettségi szintje közötti erős összefüggést, ami alátámasztja a szakirodalomban olvasottakat (Csapó és Nikolov, 2009).



5.15. ábra

*A hetedik évfolyamos diákok nemének, induktív gondolkodásuk fejlettségi szintjének, OKM matematika tesztjén elért eredményének és angol tudásszintjének előrejelző ereje a DPK teszten nyújtott teljesítményre – ítélemzés (a modellben szereplő együtthatók  $p < 0,01$  szinten szignifikánsak; IG: induktív gondolkodás; DPK: dinamikus problémamegoldó képesség; OKMm: az Országos kompetenciamérés matematika eredménye)*

A longitudinális és OKM adatokra támaszkodó elemzések is alátámasztották az induktív gondolkodás meghatározó és prediktív erővel bíró szerepét problémamegoldó képességünk fejlettségi szintjét illetően. A hét év távlatában elvégzett elemzések kiemelték a DIFER teszt és az első évfolyamos diákok számolási készségének előrejelző hatását a hetedik évfolyamos problémamegoldó képességteszten nyújtott teljesítményükre. E képességterületek első évfolyamos teljesítményei alapján 20-25%-ban magyarázható a diákok hetedik évfolyamos problémamegoldás képességteszten nyújtott teljesítménye, ami alátámasztja az erős szelekció következtében tapasztalt jelenségeket (Csapó, Molnár és Kinyó, 2009). Ugyanakkor a hatodik évfolyamos teszteredményeket integrálva a modellbe, az első évfolyamon mért területek előrejelző ereje eltűnt és egyértelműen az OKM matematika tesztjén és a longitudinális program induktív gondolkodás fejlettségét mérő tesztjén elért eredmények előrejelző ereje került előtérbe, amit a diákok neme és angol nyelvtudása követett. Utóbbi erős összefüggést mutatott a diákok induktív gondolkodásának fejlettségi szintjével ( $r=0,60$ ,  $p < 0,01$ ). Az iskolai jegyek és a diszciplináris tudásra építő tesztek egyáltalán nem jelezték előre a diákok DP képességfejlettségi szintjét. A harmadik generációs teszttel történt adatfelvétel jelentős eredménye, hogy a mérés során sikerült teljes mértékben kiküszöbölni a diákok olvasási képességének fejlettségi szintjének teljesítménybefolyásoló hatását, ami a hagyományos tesztkörnyezetben végzett képességszintvizsgálatokban jelentős hatással bírt, jelentős mértékben befolyásolta a diákok más, komplex képességterületen nyújtott teljesítményét.



#### 5.4. A dinamikus problémamegoldó környezetben alkalmazott felfedező stratégiák hatékonysága és azok változása: logfájelemzések

A disszertáció korábbi részeiben tapasztalhattuk, hogy a technológiaalapú mérés-értékelés új lehetőségeket és kihívásokat is hozott a pedagógiai empirikus kutatásokba. Számos tanulmány összegzi a számítógép-alapú mérés-értékelés fontosságát és előnyeit a hagyományos papíralapú technikákhoz képest (Csapó, Ainley, Bennett, Latour és Law, 2012). Alapvetően három fő tényező támogatja és motiválja a technológiaalapú tesztelésre való áttérést: a hagyományos, papíralapon is mért területek kapcsán tapasztalt megnövekedett mérési pontosság (pl. Csapó, Molnár és Nagy, 2014); olyan területek mérési lehetősége, amelyre hagyományos eszközökkel nem kerülhetett sor (pl.: komplex, dinamikus problémamegoldás, l. Greiff, Wüstenberg és Funke, 2012; Greiff és mtsai, 2013); végül a közvetlenül megfigyelt adatokon kívül a metaadatok rögzítése, majd a logfájlok elemzésének lehetősége (l. pl.: Tóth, Rölke, Greiff és Wüstenberg, 2014).

A pedagógiai adatbányászás (*educational data mining*, EDM), a logfájlok elemzése, hozzájárulhat ahhoz, hogy mélyebben és alaposabban megértsük és megmagyarázzuk a vizsgált jelenségek működését. Az EDM segítségével olyan kutatási kérdések megválaszolása is lehetséges, melyekre korábban, a hagyományos technikákkal nem volt lehetőség. A disszertáció e részében a DPK kutatások logfájllaira alapozva elemezzük azokat a hatékony és kevésbé hatékony problémamegoldó stratégiákat, melyeket a diákok a DPK teszt megoldása közben alkalmaztak.

A DPK teszt problémái olyan minimálisan komplex rendszerek, ahol a diákoknak azáltal, hogy manipulálhatják a bemeneti változók értékeit, amelyek, oksági kapcsolatban állnak a kimeneti változókkal, fel kell fedezniük a bemeneti és kimeneti változók közötti összefüggéseket (Wüstenberg és mtsai, 2014). Ezek a kapcsolatok a legtöbb esetben nem nyilvánvalóak, ugyanakkor a rendszer szisztematikus kontrollálásával, megfelelő felfedező és problémamegoldó stratégia alkalmazásával detektálhatóak, feltérképezhetőek.

A diákok által alkalmazott stratégiák, lépéskombinációk sokfélék, miután minden egyes bementi változónak öt értéke lehetséges és nem korlátoztuk a lépések számát (csak az időt, ami alatt a lépéskombinációkat alkalmazhatták). Ezért első lépésként szükség volt a logolt adatok átalakítására, kezelhetővé, elemezhetővé, értelmezhetővé tételére, rendszerbe foglalására.

Az egyes lépések kódolása mellett az összes, elméletileg helyes stratégiára alapozva felépítettünk egy olyan matematikai modellt, aminek segítségével elemezhetővé, áttekinthetővé váltak az alkalmazott lépéskombinációk, stratégiák. A szakirodalomból ismert úgynevezett változók kontrollja (*control of variables*, CVS; Greiff, Wüstenberg, Csapó, Demetriou, Hautamäki, Graesser és Martin, 2014) vagy egyszerre egy dolog változtatása (*very vary-one-thing-at-a-time*, VOTAT; Tschirgi, 1980; Funke, 2014) stratégián belül e modell segítségével számos további stratégia definiálását valósítottuk meg. Időelemzések segítségével teszteltük az alkalmazott stratégiák tudatosságát, elemeztük a diákok által alkalmazott stratégiák teszten belüli változását, látens profil elemzések segítségével definiáltuk a különböző típusú problémamegoldókat, valamint összevetettük a különböző felfedező stratégiák kapcsolatát a tudás elsajátítás képességét mérő itemeken nyújtott teljesítményekkel, valamint az általános problémamegoldó képességgel. Végül elemeztük, hogy a probléma megoldásával töltött idő és kattintások száma hogyan függ össze a teljesítményekkel.

Összefoglalóan az itt bemutatott elemzések egyrészt bemutatják ezt az új elméleti, az elemzéseket lehetővé tevő modellt, másrészt empirikus adatok segítségével hozzájárulnak a dinamikus problémamegoldó környezetben alkalmazott és alkalmazható stratégiák megértéséhez. A logfájl elemzések kitágították a teszt problémáinak megoldása során alkalmazott fázisok körét. Míg a diákok válaszai alapján a modellrajzolás és a rendszer működtetése (modellalkalmazás) fázisokat tudtuk elkülöníteni, addig a logfájlok segítségével a modellrajzolás fázisát megelőző felfedező fázis is számszerűsíthetővé vált. Ezen új első fázisban a diákok szabadon manipulálhatták a bemeneti változók értékét, hogy megfigyelhessék azok hatását a kimeneti változók értékére. A bemeneti és a kimeneti változók közötti kapcsolat vagy a rendszer sajátdinamikája a bemeneti változók megfelelő, szisztematikus változtatásával, helyes problémamegoldó stratégia alkalmazásával fedezhető fel. E felfedező, új tudás elsajátításának hatékonyságát mérő fázissal párhuzamosan a diákoknak a felismert összefüggéseket fel kellett rajzolniuk, vizualizálniuk kellett egy modellben, majd a problémamegoldás harmadik fázisában a korábban megismerteknek megfelelően működtetniük a rendszert bizonyos előre meghatározott célok elérése érdekében (tudás alkalmazása fázis).

A szakirodalomban elérhető elemzések a DPK teszt problémáinak második és a harmadik megoldási fázisában nyújtott teljesítményeken alapulnak, és azt kötik közvetlenül a tudás elsajátítása és tudás alkalmazása faktorokhoz (Greiff és mtsai, 2013; Wüstenberg, Greiff, Molnár és Funke, 2014). A logfájl elemzések segítségével azonban közvetlenül feltérképezhetőek az első fázisban alkalmazott stratégiák.

#### 5.4.1. A MicroDYN megközelítés teljes, alap és minimális stratégia modellje

A MicroDYN problémák alapvetően kevés számú változót és oksági kapcsolatot tartalmaznak. A minimálisan komplex rendszerre épülő problémák megoldásához használható jó stratégiák leírhatóak egy véges modellben. A modell felépítéséhez a felfedező fázisban végzett minden egyes tevékenységet rögzítettünk, beazonosítottunk, majd különböző szempontok szerint klasztereztünk. A klaszterezés eredményeképpen háromféle stratégiát definiáltunk: teljes stratégia, alap stratégia és minimális stratégia.

A teljes stratégia magában foglalja mindazt, amit a tesztet megoldó személy pontosan tett a probléma megoldásának első fázisában. Tartalmazza az összes – bemeneti változók manipulálását lehetővé tevő – csúszkával és – azok hatását a kimeneti változókon megjelenítő – alkalmazás gombbal történő interakciót abban a sorrendben, ahogy azok történtek a problémamegoldás első fázisában.

Az alapstratégia a teljes stratégia egy része. Az alapstratégia felírása során továbbra is figyelembe vesszük az időfaktort, de ez a stratégia már csak azokat a tevékenységeket tartalmazza, aminek segítségével a diák új információhoz jutott a rendszer megismerése, felfedezése kapcsán. Ez azt jelenti, hogy a teljes stratégiához képest a következő tevékenységeket nem tartalmazza az alapstratégia, ezek a tevékenységek az alapstratégia generálása során a teljes stratégiából törlésre kerültek:

- ha a feladaton belül a korábbiakkal azonos bemeneti változó beállítás (csúszkabeállítás) történt,
- ha egyszerre több mint egy nem ismert működésű, hatású bemeneti változó került változtatásra,

- ha, bár a bemeneti változók új, korábban még nem alkalmazott beállítási állapota történt meg, de a beállítás során használt bemeneti változók hatása, működése korábbi beállításokból már ismert lehetett.

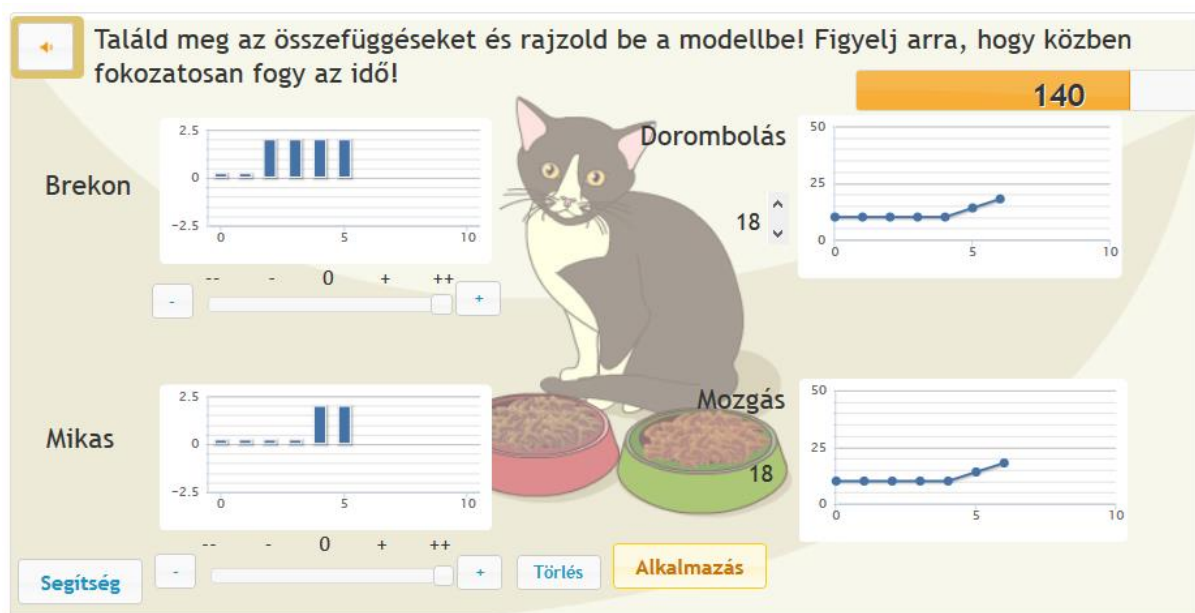
Végül az alapstratégiából a teljes stratégia figyelembe vételével a minimális stratégia generálása valósult meg. A minimális stratégia időfaktor nélkül azokat a tevékenységeket foglalja magába, ahol a diák új információt tudott kinyerni a rendszer működése kapcsán és azt a legideálisabb lépéskombinációval tette meg (ennek fontosságát l. később). Ha a diák által alkalmazott stratégiában megjelent az általunk keresett valamely, elegendő információt szolgáltató manipuláció, akkor függetlenül a többi adott problémán belül alkalmazott lépéstől, hozzárendeltük az adott stratégiát.

Mindhárom stratégián belül – ahol értelmezhető – az egyes tevékenységek, lépések kódolása (nem pontozása) a következők szerint valósult meg (az állapotok rögzítése az alkalmazás gomb lenyomásához kötődik minden esetben):

- 1) Egyetlen egy bemeneti változó értékének változtatása történt meg, amíg a többi bemeneti változó értéke semleges értéken (pl.: nullán) maradt. Ezen kombináció kódolása +1-el valósult meg.
- 2) Egyetlen egy bemeneti változó változtatása történt meg, a többi bemeneti változó értéke nem semleges állapotban, de egy korábban már kipróbált állapotban volt. E szcenáriót +2-vel kódoltuk.
- 3) Egyetlen egy bemeneti változó változtatása történt meg, a többi bemeneti változó értéke nem semleges és nem is egy korábban már alkalmazott kombinációban volt, viszont hatásuk a korábbi lépések eredményeként ismert lehet. Ezt a féle lépést +3-al kódoltuk.
- 4) Minden egyes bemeneti változó értéke semleges (pl.: nulla) értéken maradt (különösen lényeges a sajátdinamikával rendelkező rendszerek kiismerése során). Ezt a féle kombinációt +A-val jelöltük.
- 5) Több mint egy bemeneti változó értékének egyidejű megváltoztatása, ugyanakkor az alkalmazott kombináció nem szolgáltat plusz információval a rendszer működése kapcsán. Ezt a féle lépést -X-el jelöltük.
- 6) Ugyanaz a kombináció került beállításra, ami a feladaton belül már egyszer megtörtént. A feladatot megoldó személy így nem jutott plusz információhoz a rendszer viselkedése kapcsán. A -0 jelölést alkalmaztuk ebben az esetben.
- 7) A bemeneti változók beállításának új kombinációja került alkalmazása, ugyanakkor a beállításban használt bemeneti változók hatása már korábbi beállításokból ismert, nem jelentett plusz információt e szcenárió futtatása. Ezt a féle beállítást +0-val jelöltük.
- 8) Több ismeretlen működésű bemeneti változó értéke került egyszerre változtatásra, ugyanakkor a korábbiak és e kombináció hatása alapján elméletileg, lineáris egyenletrendszer megoldása segítségével lehetséges a bementi változók működésének kiszámolása. Ezt a féle lépést +4-el jelöltük.
- 9) Az alapstratégia kapcsán egy extra +5-ös kódot is bevezettünk arra az esetre, amikor az utoljára alkalmazott szcenárió hatása alapján lehetséges volt az összes bemeneti változó működésének kiszámolása. Ez a féle lépés a minimális stratégiának nem eleme.

A szakirodalomból ismert VOTAT-stratégiák (Fischer, Greiff és Funke, 2012) közé sorolható +1, +2 és +3 jelölésű kombináció (esetlegesen a +A-val jelölt is), míg a -X, -0, +0, +4 és +5-ös stratégiákkal egyáltalán nem foglalkoztak a korábbi elemzések. A következőben két példán keresztül szemléltetjük a kódolás folyamatát, illetve a minimális stratégia alap és teljes stratégiából való generálásának folyamatát.

Az 5.16. ábra egy két bemeneti és két kimeneti változóval rendelkező példát mutat. A rendszer feltérképezése során az első két lépésben mindkét bemeneti változó értékét nullán tartottuk, aminek hatására nem változtak a kimeneti változók értékei. A harmadik és negyedik lépésben a brekon nevű bemeneti változó értékét 2-re állítottuk, míg a mikas nevű továbbra is nullán maradt. Ez a féle változtatás sem gyakorolt hatást a kimeneti változók értékére. Ötödik és hatodik lépésben a brekon nevű bemeneti változó értéke továbbra is 2-es értéken maradt, de ehhez még hozzáadódott a mikas nevű bemeneti változó értékének 2-re állítása. Ennek hatására mindkét kimeneti változó értéke azonos mértékben nőni kezdett.



5.16. ábra

*A MicroDYN problémák első fázisának feltérképezése (2 bementi, 2 kimeneti változó)*

E típusú problémamegoldó stratégia alkalmazása során a teljes stratégia kódolása a következőképpen alakul: +A, -0, +1, -0, +2, -0. Miután a második, negyedik és hatodik lépés a korábbi kombinációk ismétlése volt, ezért -0 kóddal illettük. Az első lépésben a sajátdinamika felfedezéséhez nélkülözhetetlen (nulla-nulla) beállítás került alkalmazásra (+A). A harmadik lépés a VOTAT-stratégia legtisztább alkalmazása (egyszerre csak egy bementeti változó értékének módosítása, míg a többi bementeti változó értékének semleges szinten tartása, +1), míg az ötödik lépés során alkalmazott stratégia is VOTAT-stratégia, hisz a negyedik lépéshez képest csak egy bementeti változó értéke változott, mégsem beszélhetünk ugyanarról a stratégiáról, mint a harmadik lépésben tettük (+2). Az ötödik lépés után minden szükséges információ már a problémamegoldó rendelkezésére állt.

Ugyanezen lépéssorozat alapstratégiája: +A, +1, +2, miután a problémamegoldó többi lépése nem vezetett még nem ismert információ kinyeréséhez. Az időfaktorfüggetlen minimális

stratégia ebben az esetben szintén +A, +1, +2. E lépések kapcsán tudott a tesztelt személy új információhoz jutni a rendszer működése kapcsán.

Az 5.17. ábra egy két bemeneti és egy kimeneti változóval rendelkező rendszer feltérképezését mutatja. A diák által alkalmazott teljes stratégia a következő kódolással írható le: +1+2+0+0+0+0-0-0. A második lépés után gyakorlatilag az összes információ a rendelkezésére állt, ami szükséges a modell felrajzolásához. Ugyanezen teljes stratégia alapstratégiája: +1+2, minimális stratégiája viszont +1+1, miután a lépéssorozat első és hatodik lépésében a legtisztább VOTAT-stratégia alkalmazására került sor (egyik változó nullától különböző értéken, a másik változó semleges értéken).



5.17. ábra

*A MicroDYN problémák első fázisának feltérképezése (2 bementi, 1 kimeneti változó)*

A MikroDYN problémák első fázisára kidolgozott teljes, alap és minimális stratégia modell, illetve az egyes lépéskombinációk kódolása minden egyes hasonló, minimális komplex rendszerre építő problémára alkalmazható. A rendszer három különböző VOTAT-stratégiát különböztet meg, mely a leghatékonyabb stratégia a MicroDYN problémákhoz hasonló rendszerek kiismerése során (Fischer és mtsai, 2012). A VOTAT-stratégiákban közös, hogy a probléma megoldója szisztematikusan egyszerre mindig csak egy bemeneti változó értékét változtatja meg, ezzel könnyen felismerhetővé teszi a módosított bemeneti változó hatását a kimeneti változókra. A szakirodalomban leggyakrabban alkalmazott VOTAT-stratégia, amikor a változtatott bemeneti változón kívül a többi bemeneti változó értéke semleges, nulla szinten van (elszigetelt változók stratégiája). A disszertáció itt bemutatott elemzéseiben három különböző VOTAT-stratégiát, valamint számos, helyes megoldáshoz vezető nem VOTAT-stratégiát is azonosítottunk. Alkalmazásuk sikerességét összevetettük a diákok teljesítményével. Előzetes hipotéziseink szerint a „legtisztább”, a változók teljes elszigetelésén (Wüstenberg, Stadler, Hautamäki és Greiff, 2014) alapuló VOTAT-stratégia alkalmazása nagyobb valószínűség mellett vezet helyes megoldáshoz, mint más VOTAT vagy nem VOTAT-stratégia használata.

#### 5.4.2. A logfájlelemzések céljai

A logfájlelemzések célja a MicroDYN problémák során alkalmazott explorációs stratégiák feltérképezése és a kidolgozott modell hatékonyságának, általánosíthatóságának bizonyítása. Annak meghatározása, hogy (1) milyen arányban vezet egy helyes stratégia alkalmazása magas teljesítményhez; (2) milyen arányban vezet a VOTAT-stratégia alkalmazása magas teljesítményhez, melyik a leghatékonyabb VOTAT-stratégia; (3) a stratégiahasználat tudatossága mennyiben határozza meg a DPK teszten nyújtott teljesítményt; (4) azonosíthatók-e különböző típusú problémamegoldók, arányuk hogyan változik az életkor előrehaladtával; (5) hogyan változnak az alkalmazott explorációs stratégiák a teszt megoldása közben, (6) milyen mértékben befolyásolja a problémamegoldó által tesztelt kombinációk száma és a problémák feltérképezésével töltött idő a problémák második és harmadik fázisában nyújtott teljesítményeket.

#### 5.4.3. Módszerek

Az elemzéseket a két 2014-es és a 2015-ös egyetemisták körében történt adatfelvétel eredményeivel ( $n=1259$ ) és logadataival végeztük. 2014-ben egyrészt 3–12. évfolyamos tanulók részvételével történt kutatás, melynek 6. ( $n=677$ ), 7. ( $n=607$ ) és 8. ( $n=942$ ) évfolyamos részmintáját vontuk be az elemzésekbe, másrészt a longitudinális adatfelvétel 7. évfolyamos ( $n=2237$ ) adataira alapoztuk az e részben ismertetett eredményeket.

A kutatás és mintaválasztás oka egyrészt a korábbi elemzések alapján a fejlődés szempontjából szenzitív és gyorsan változó 6-8. évfolyamra eső korszak választása, a hasonló korosztályban történt 2013-as adatfelvételhez képest nagyobb mintaelemszám, továbbá lehetőség az évfolyamok közötti összehasonlításra volt. Miután a 6-8. évfolyamos részmintán – a kutatás felépítése következtében – ugyanazon DPK teszt került kiközvetítésre, ezért az adott kutatáson belül nemcsak a problémák, hanem a problémák teszten belüli pozíciója, elhelyezkedése is azonos volt.

A két 2014-es adatfelvétel közel azonos kohortjának kiközvetített tesztje nem teljesen volt azonos. A longitudinális adatfelvételben 90 másodperc, míg a 6-8-os diákoknak kiközvetített tesztben 180 másodperc állt rendelkezésre a probléma feltérképezéséhez (l. korábban). Az előbbiben nem, az utóbbiban meghallgathatták a diákok az instrukciókat. A kizárólag 7. évfolyamon kiközvetített tesztben 9, míg a tágabb életkori intervallumon megoldott tesztben 10 probléma szerepelt. Miután az egyetemista mintán felvett és a 7. évfolyamos longitudinális kutatás keretein belül felvett adatok felhasználására a profilképzés során került csak sor és a részletes stratégiaelemzést a 6-8. évfolyamosok logadataira építjük, ezért a továbbiakban az ő tesztjük felépítésére fókuszálunk.

Az 5.3. táblázat a bemeneti és kimeneti változók száma szerint összefoglalja a 6-8. évfolyamos diákok tesztjében szereplő problémák komplexitását, a változók közötti lehetséges és meglévő kapcsolatok számát, valamint az esetleges sajátdinamika meglétét. A tesztben lévő problémák az összes lehetséges típusú hatást tartalmazták, amit egy maximum három bementi és három kimeneti változóból álló rendszer tartalmazhat, mindegyiket különböző és nem valós történetbe ágyazva. A kutatás mintájának és mérőeszközének részletes leírását l. 4.2.2. részben.

A problémamegoldók látens osztályaink meghatározása érdekében mindkét mintán látens profil (*Collins és Lanza, 2010*) elemzést végeztünk. A látens profil elemzés a

problémamegoldók olyan típusait keresi, akik az alkalmazott problémamegoldó stratégiák tekintetében hasonló mintázatot mutatnak. A látens csoportok számának meghatározása során több kritériumot is alkalmaztunk: relatív illeszkedésindexeket, mint az AIC (*Akaike Information Criterion*, Akaike Információs Kritérium), a BIC (*Bayesian Information Criterion*, Bayesi Információs Kritérium) és az aBIC (*adjusted Bayesian Information Criterion*, korrigált Bayesi Információs Kritérium) mutatókat. Mindhárom esetben az alacsonyabb értékek a jobban illeszkedő modellt jellemzik. Az entrópia segítségével megállapítható, hogy milyen pontosan tudjuk a személyeket a megfelelő csoportokba sorolni, kategóriákhoz hozzárendelni, azaz mennyire homogének a csoportok. A magasabb értékek a jobb illeszkedést jelzik. Végül a Lo-Mendell-Rubin valószínűségi hányados teszttel (*Lo-Mendell-Rubin Adjusted Likelihood Ratio Test*) összehasonlítottuk az  $n$  számú látens osztályt tartalmazó modellt, az  $n-1$  számúval. A szignifikáns  $p$ -érték ( $p < 0,05$ ) azt jelzi, hogy az  $n-1$  számú modell elvetésre kerül az  $n$  csoportot tartalmazó modell javára, ugyanis az aktuálisan tesztelt modell jobban illeszkedik, mint az azt megelőző (*Muthén és Muthén*, 2012).

### 5.3. táblázat. A 6-8. évfolyamosok tesztjében szereplő problémák minimális komplexitása

A probléma tesztben elfoglalt helye	Bementi változók száma	Kimeneti változók száma	Kapcsolatok száma	Saját- dinamika	A különböző beállítások optimális száma, amivel a rendszer felfedezhető
1	2	2	2	0	2
2	2	1	2	0	2
3	2	2	2	0	2
4	2	2	2	0	2
5	3	2	3	0	3
6	3	3	3	0	3
7	3	2	2	1	4
8	3	3	4	0	4
9	3	2	3	1	4
10	3	3	4	1	4

A látens profil elemzés kategorikus látens változóit a logfájlok alapján generált 0,1,2 kategóriákat tartalmazó változók adták. A problémák kapcsán minden egyes diákhoz hozzárendeltünk még egy kategoriális változót. A változó azt jellemezte, hogy az adott diák az adott probléma kapcsán mennyire alkalmazta a teljes mértékben elszigetelt változókezelésre építő VOTAT-stratégiát. Ha egyáltalán nem tette, akkor az új változó értéke 0. Ha részlegesen tette, azaz például nem mind a három bemeneti változó kapcsán, hanem csak kettő vagy egy változóra alkalmazva, akkor 1-el kódoltuk az új változó értékét. Ha a problémában szereplő összes bemeneti változóra alkalmazta, akkor pedig 2-es kódolást kapott az új változó.

#### 5.4.4. Eredmények

A teszt hagyományos változói (a diákok válaszai) szerint számolt reliabilitásmutatója az érintett évfolyamokon – a problémamegoldás két fázisát együtt (20 item) elemezve –  $\alpha = 0,80$ . A teszt modellépítés résztesztjének (10 item) reliabilitásmutatója  $\alpha = 0,72$ . A tudás elsajátítás hatékonyságának fázisát mérő részteszt megbízhatósági mutatója jelentős mértékben

megnövekedett, ha azt a diákok teszten nyújtott, logfájlok segítségével kinyert viselkedése alapján számoltuk. A korábban részletezett kódolás alapján, ha a diák által alkalmazott stratégiával meg lehetett oldani a problémát, akkor 1 pontot, ha nem, akkor 0 pontot kapott. Ezzel a módszerre minden egyes problémához hozzárendeltünk diákonként még egy változót, azaz teszt szintjén 10 új itemet. A generált, logfájlokon alapuló új változó segítségével számolt megbízhatósági mutató értéke (10 item):  $\alpha=0,91$ , azaz a diákok felfedező stratégiáinak direkt kódolása megbízhatóbban jellemzi a diákok tudás elsajátítása terén mutatott képességszintjét, mint a tanultak modellépítés formájában történő leképezése. Évfolyamonkénti bontásban  $\alpha_{\text{evf6}}=0,91$ ,  $\alpha_{\text{evf7}}=0,92$ ,  $\alpha_{\text{evf8}}=0,91$ . A 30 itemet együtt kezelve, mint a problémamegoldás három fázisának résztesztjeit, a megbízhatósági mutató értéke a teljes mintán  $\alpha=0,88$  ( $\alpha_{\text{evf6}}=0,88$ ,  $\alpha_{\text{evf7}}=0,89$ ,  $\alpha_{\text{evf8}}=0,90$ ).

Az elemzések alapján megállapítható, hogy a felfedezés során alkalmazott jó stratégia nem minden esetben vezetett magas teljesítményhez (l. 5.4. táblázat) és fordítva, nem csak a helyes stratégia alkalmazása eredményezett magas teljesítményt. A legalacsonyabb komplexitású problémák esetén (2 bemeneti, 1 kimeneti változó). A diákok közel háromnegyede helyes stratégiát alkalmazott a rendszer kiismerése során, de csak a diákok fele tudta helyesen leképezni a megszerzett tudást és jól felépíteni a rendszer működését ábrázoló modellt. Ahogy nőtt a problémák komplexitása, úgy csökkent a minden szükséges információ kinyerését megvalósító stratégia alkalmazásának aránya. A sajátdinamikával rendelkező problémák esetén jelentős mértékben csökkent a megfelelő stratégiát alkalmazó diákok köre és nekik is csak egy kis hányada, mintegy hatoda hozta meg végül a helyes döntést.

5.4. táblázat. Az elméletileg helyes vagy a rendszer kiismeréséhez nem elegendő stratégia alkalmazásának sikeressége a problémák komplexitásának fényében (minimális stratégia)

Probléma komplexitása	Gyakoriság (%)					
	Elméletileg helyes stratégia			Nem helyes stratégia		
	0	1	Össz.	0	1	Össz.
2-1	20,30	49,25	69,55	8,27	22,18	30,45
2-2	32,00	35,10	67,10	23,89	9,01	32,90
3-2	22,84	24,61	47,45	44,72	7,83	52,55
3-3	24,17	26,84	51,01	37,02	11,97	48,99
3-2 sajátdinamika	6,36	1,20	7,56	92,01	0,43	92,44
3-3 sajátdinamika	3,48	1,68	5,16	93,82	1,02	94,84

Megj.: A pozíciós hatás kizárása érdekében a táblázatban szereplő adatok a tesztben előforduló első adott komplexitású problémára vonatkoznak.

A nem elegendő információ kinyerését adó stratégiát alkalmazó diákok között is relatív nagy számban voltak azok, akik végül helyes modellt építettek fel, miközben az alkalmazott stratégia alapján a modell teljes felépítéséhez nem rendelkeztek az összes információval. A legalacsonyabb komplexitású rendszereknél igen magas volt a találgatás aránya, a helytelen stratégiát alkalmazók mintegy kétharmada helyes modellt rajzolt fel (körülbelül annyian, ahányan a helyes stratégiahasználat ellenére is rosszul építették fel a modellt). Ez az arány jelentős mértékben csökkent bonyolultabb rendszerek alkalmazása során. A három bemeneti,



két kimeneti változóval rendelkező probléma esetén például már csak a helytelen stratégiát alkalmazók hatoda hozott végül intuitíven helyes döntést.

A fejlődési, változási tendenciák detektálása végett összevetettük a 6. és a 8. évfolyamos diákok válaszait, illetve helyes és helytelen stratégiahasználatát (5.5. táblázat). A korábbi eredményekre alapozó előzetes hipotézisünk szerint minden egyes problémátípus kapcsán fejlődést vártunk. Ennek ellenére nagyon kismértékű fejlődés történt a helyes stratégiahasználat tekintetében 6. és 8. évfolyam között. A helyes stratégiát alkalmazó diákok 8. évfolyamon sikeresebben képezték le a kinyert információkat, mint 6. évfolyamon, azaz nagyobb arányban építették fel jól a rendszer szerkezetét reprezentáló modellt.

5.5. táblázat. Az elméletileg helyes vagy a rendszer kiismeréséhez nem elegendő stratégia alkalmazásának sikeressége a problémák komplexitásának fényében 6. és 8. évfolyamon (minimális stratégia)

Évf.	Probl. kompl.	Gyakoriság (%)						t (strat.)	t (mod.)
		Elméletileg helyes stratégia			Nem helyes stratégia				
		0	1	Össz.	0	1	Össz.		
6.	2-1	19,91	49,04	68,96	10,24	20,79	31,03	n.s.	n.s.
8.		21,14	52,39	73,53	6,37	20,08	26,46		
6.	2-2	36,25	31,87	68,12	22,66	9,21	31,87	n.s.	-3,81**
8.		29,21	39,97	69,18	21,32	9,48	30,81		
6.	3-2	22,59	22,74	45,34	46,52	8,12	54,65	-2,07*	-1,90*
8.		22,58	28,12	50,70	41,36	7,92	49,29		
6.	3-3	27,34	23,03	50,37	39,07	10,55	49,62	n.s.	-4,09**
8.		22,23	32,27	54,51	33,40	12,07	45,48		
6.	3-2	5,02	0,41	7,78	5,83	0,14	5,98	-2,16*	-3,30**
8.	saját-di- namika	8,44	2,59	11,03	4,84	0,11	4,95		
6.	3-3	3,46	1,20	4,67	81,74	13,57	95,32	n.s.	-2,91**
8.	saját-di- namika	2,60	5,54	8,15	74,06	17,78	91,84		

Megjegyzés: \*:  $p < 0,05$ , \*\*:  $p < 0,01$  szinten szignifikáns, n.s.: nem szignifikáns.

Összességében megfogalmazható, hogy a diákok által alkalmazott stratégia hatékonysága nem minden esetben egyezett meg teljesítményük minőségével. A megoldáshoz szükséges összes információ kinyerését biztosító stratégia és a rendszer működését pontosan leíró modellek felépítésének aránya változó volt. A legalacsonyabb komplexitású problémák esetén a helyes stratégiát alkalmazók kétharmada reprezentálta jól a probléma szerkezetét, míg a sajátdinamikával nem rendelkező, de már három bemeneti változót tartalmazó problémák esetén ugyanezen diákok már csak fele tudta meghozni a helyes döntést. Ha a felfedezendő rendszer sajátdinamikát is tartalmazott, akkor a legmagasabb képességszintű, helyes stratégiát alkalmazó diákoknak (a minta 5-8%-a) is csak ötöde, azaz a diákok 1-2%-a építette fel az adott probléma működését reprezentáló modellt jól és oldotta meg ezzel az adott problémát.

Elemeztük, hogy mi az aránya az alkalmazott, minden szükséges információ kinyerését biztosító stratégiákon belül a VOTAT és nem VOTAT-stratégia használatának. A VOTAT-stratégiák fő jellemzője, hogy két, plusz információ kinyerését célzó lépés között minden esetben csak egyetlen egy változó manipulációja történik. A két bemeneti változót (és sajátdinamikát nem) tartalmazó problémák esetén a korábban említett kódolást alkalmazva a

VOTAT-stratégiák a következő minimális stratégiákkal írhatóak le teljes körűen: +1+1, +1+2, +1+3, míg a +4-es stratégia már a nem VOTAT stratégiák közé sorolandó. Három bemeneti változó esetén a VOTAT-stratégiák: +1+1+1; +1+1+2; +1+1+3; +1+2+2, +1+2+3, +1+3+2; +1+3+3 és nem VOTAT-stratégia: +1+4.

A VOTAT és nem VOTAT, de elegendő információt biztosító stratégiák alkalmazásának aránya változik a problémák komplexitásával (5.6. táblázat). A két bementi változót tartalmazó rendszerekben kb. 87/13 az arány, míg a három bemeneti változóval rendelkező problémák esetén már 80/20. Ez arra utal, hogy a stratégiaelemzések során nem szabad figyelmen kívül hagyni a nem VOTAT, de helyes stratégiát alkalmazókat sem, akik a – kizárólagosan VOTAT-stratégiákra fókuszáló – szakirodalomban általában a helytelen stratégiahasználók közé soroltak. Az további kutatást igényel, hogy az egyes kultúrák, nemzetek között változik-e ez az arány, miután a PISA problémamegoldás adatait elemezve (Greiff, 2015) hazánkban kimagaslóan magas a helytelen stratégiát alkalmazó, de helyesen válaszoló diákok aránya.

A PISA elemzésekben kizárólagosan a VOTAT, illetve a legegyszerűbb (egy bemeneti változó nem nulla pozícióban, a többi bementi változó nulla, semleges pozícióban) VOTAT-stratégia kezelésére került csak sor. A többi stratégia, így a nem VOTAT, de helyes stratégiák is a helytelen stratégiák közé sorolódtak, miközben a VOTAT-stratégiák 50-60%-a volt csak elszigetelt változókezelésen alapuló stratégia (5.6. táblázat).

*5.6. táblázat. Az elméletileg helyes stratégiákon belül a VOTAT, a VOTAT stratégiákon belül az elszigetelt változókezelést alkalmazó VOTAT, valamint a nem VOTAT stratégiák aránya*

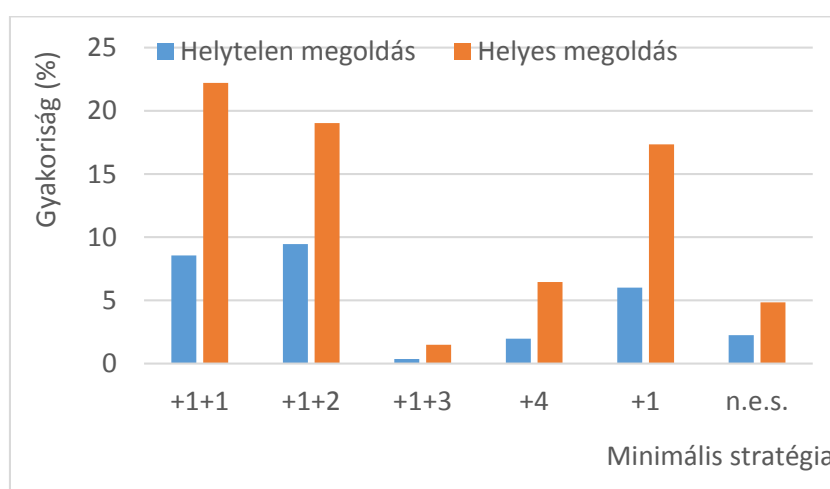
Bemeneti/kimeneti változók száma	Elméletileg helyes stratégia (gyakoriság, %)	
	VOTAT (elszigetelt változókezelés VOTAT %-ban)	nem VOTAT
2-1	87,91 (50,40)	12,09
2-2	86,97 (51,69)	13,02
3-2	78,71 (54,20)	21,28
3-3	80,21 (57,39)	19,78

A VOTAT stratégiahasználaton belül felmerülő kutatási kérdés, hogy vajon azonosan hatékony-e az összes VOTAT-stratégia, van-e olyan VOTAT-stratégia, ami alkalmazása mellett a diákok nagyobb valószínűséggel építik fel helyesen a rendszer működését leíró modellt. Az elemzés során a teljes és alapstratégiából generált minimális stratégiákkal dolgoztunk, ahol figyelmen kívül hagytuk az ismételt beállításokat és a sorrendiséget, az időfaktort. Minden egyes problémastruktúrára (bementi és kimeneti változók száma szerint, figyelmen kívül hagyva a kapcsolatok számát) lefuttattuk az elemzéseket. Minden esetben a tesztben előforduló első, adott struktúrával rendelkező probléma megoldása során logolt adatokra építettünk. Az eredményeket ábrázoló diagramokon megjelenítettük a nem elegendő stratégiahasználat (n.e.s.) és a helyes és helytelen válaszok arányát is.

A legegyszerűbb, kettő bemeneti változót és egy kimeneti változót tartalmazó feladatoknál a korábbi kódolást alkalmazva VOTAT-stratégiának számít: +1+1; +1+2; +1+3, míg a nem VOTAT, de helyes stratégiák közé sorolhatóak: +4. A teljes rendszer kiismerése vonatkozásában a nem elegendő stratégiákhoz tartozik a +1 stratégia, mégis külön kezeltük,

miután a kis számú bemeneti változó miatt már egy bementi változó viselkedésének ismerete alapján nagy valószínűséggel felrajzolható a modell működése.

Ezen problémák esetén, miután relatív kevés változót és összefüggési lehetőséget tartalmazott a rendszer (3 db: csak az első bementi változó hat a kimeneti változóra, csak a második bementi változó hat a kimeneti változóra, minkét bemeneti változó hat a kimeneti változóra, a negyedik esetet alapbeállításban kizártuk, amikor egyik bementi változó sem hat a kimeneti változóra), nemcsak a helyes stratégia alkalmazása esetén, de a nem elegendő stratégiahasználat mellett is magas volt a helyes válaszok aránya, minden esetben magasabb, mint a helytelen válaszoké. A VOTAT-stratégiák közül a +1+1, illetve a +1+2 típusú került leggyakrabban alkalmazásra (5.18. ábra; az elemzések során például a +A+1+1 típusú stratégiát a +1+1 lépéskombinációhoz soroltuk abban az esetben, ahol a probléma feltérképezéséhez nem volt szükség sajátdinamika vizsgálatára).



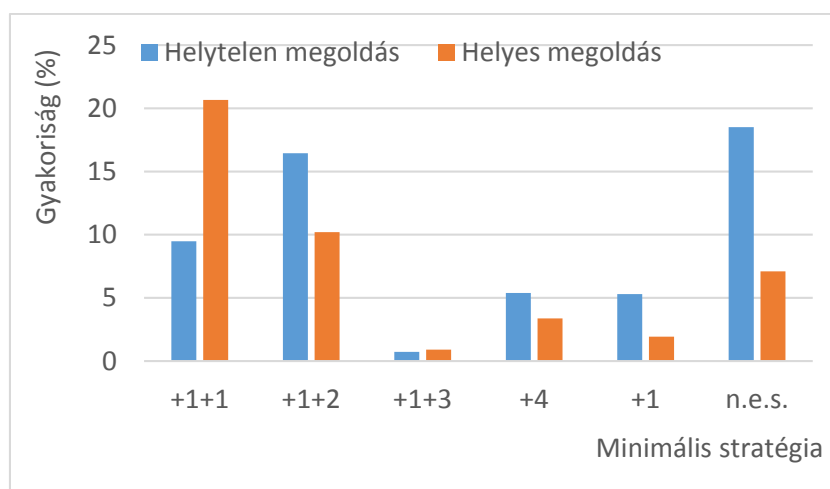
5.18. ábra

*A legegyszerűbb, kettő bemeneti és egy kimeneti változót tartalmazó problémák feltérképezése során alkalmazott minimális stratégiák a teljesítmény függvényében (n.e.s.: a szükséges információk kinyeréséhez nem elegendő stratégia; stratégiakódoláshoz l. 5.4.1. részt)*

Arányaiban a +1+1 stratégiát alkalmazók – azaz amikor első lépésben megnézték az első bemeneti változó hatását, a másik bemeneti változó értékét nullán tartva, majd megnézték a második bemeneti változó hatását, az első változó értékét nullán tartva – voltak a legsikeresebbek, háromnegyedük helyesen rajzolta fel a modellt. A +1+2 stratégiát alkalmazók (megnézték az egyik bemeneti változó hatását, majd az értéket megtartva megváltoztatták a másik bemeneti változó értékét) kétharmada tudta helyesen felrajzolni a modellt. A kicsit alacsonyabb megoldottsági ráta oka lehet, hogy ők az eredmény értelmezése során figyelmen kívül hagyták azt, hogy a +1+2 stratégia alkalmazásával a második lépésben már kumulált hatást látnak és nem a másodszorra változtatott bemeneti változó elszigetelt hatását. A +1+3 VOTAT-stratégiát alkalmazók is, bár számuk alacsony volt, nagyobb arányban oldották meg helyesen a problémát, mint helytelenül. Hasonló jelenséget tapasztaltunk a +4 helyes, de nem VOTAT-stratégiát használó diákok körében is, kétharmaduk helyes döntést hozott a modell felépítése kapcsán. A nem helyes stratégiahasználók között túlerepresentáltak a +1 lépéskombinációt alkalmazók, akik nagy része a rendszer egyszerű felépítése következtében

ezzel a stratégiával is sikeres problémamegoldónak bizonyultak. Szintén ez okozhatta azt, hogy az egyéb, nem elegendő stratégiát alkalmazók is intuitíve nagyon jól teljesítettek az egyszerű szerkezetű problémákon.

A stratégiahasználat helyessége és a kinyert információk alapján felállított modell helyessége közötti összefüggés jelentős mértékben megváltozott a kimeneti változók tekintetében eggyel nagyobb fokú komplexitással rendelkező problémák esetén (5.19. ábra). A kettő bemeneti és kettő kimeneti változóval rendelkező rendszerek feltérképezése során alkalmazható VOTAT és helyes nem VOTAT-stratégiák köre nem változott a két bemeneti és egy kimeneti változóval rendelkező problémákhoz képest, viszont, miután nőtt a lehetséges kapcsolatok és ezért a felállítható modellek száma, jelentős mértékben csökkent az intuitíven jó problémamegoldók aránya.

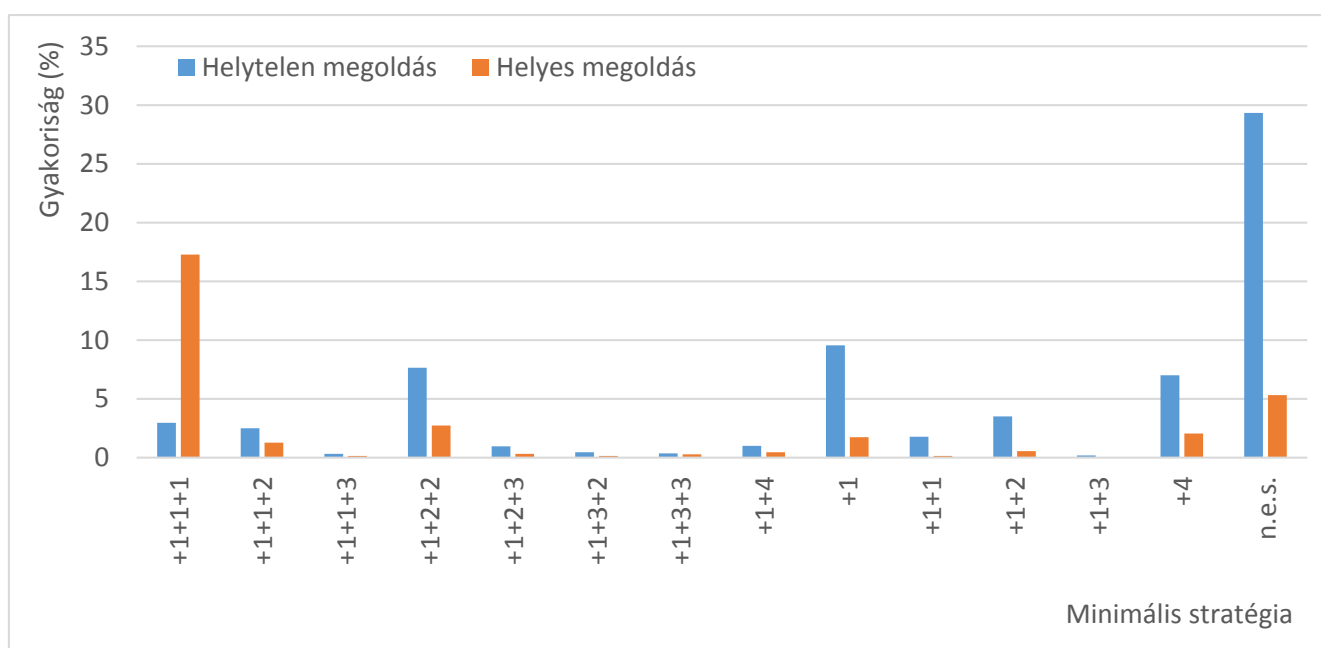


5.19. ábra

*A két bemeneti és két kimeneti változót tartalmazó problémák feltérképezése során alkalmazott minimális stratégiák a teljesítmény függvényében (n.e.s.: a szükséges információk kinyeréséhez nem elegendő stratégia; stratégiakódoláshoz l. 5.4.1. részt)*

Továbbra is hatékony problémamegoldó stratégiának bizonyultak a VOTAT-stratégiák, azonban markánsabban kirajzolódott az elszigetelt változók kezelésére alapozó stratégia előnye (Wüstenberg, Stadler, Hautamäki és Greiff, 2014) a többi VOTAT-stratégia fényében, amikor egymástól függetlenül kezelték és ellenőrizték a bemeneti változók kimeneti változókra gyakorolt hatását a diákok. A +1+1 stratégiát alkalmazók kétharmada helyes modellt állított fel a problémamegoldás modellépítés fázisában, míg a +1+2 stratégiát használók többsége helytelen. Feltételezhetően ők figyelmen kívül hagyták azt, hogy az alkalmazott stratégia nem elszigetelten, hanem kumuláltan nézi a bemeneti változók hatását. A +1+2 stratégiát alkalmazók kettőötöde helyes, míg háromötöde helytelen modellt állított fel, holott a rendszer működésének pontos leírásához szükséges összes információ rendelkezésükre állt. Továbbra is kis arányban fordultak elő a +1+3, illetve +4 stratégiát működtető problémamegoldók. Ennek ellenére a nem VOTAT, de helyes +4-es stratégiát alkalmazó jó problémamegoldók ismét felhívták a figyelmet arra, hogy a MicroDYN rendszerekben történő problémamegoldó stratégiavizsgálatok során nem elegendő a VOTAT-stratégiákra szűkíteni az elemzéseket. Az alkalmazott stratégia tekintetében a helytelen stratégiahasználók közé sorolható diákok harmada oldotta meg jól intuitíven az ilyen típusú problémákat, azaz már kettő kimeneti változó

alkalmazása esetén jelentős mértékben visszaszorítható a helyes találgatások aránya. A három bemeneti és két kimeneti változóval rendelkező rendszerek még kisebb teret adtak a találgatásnak, amit az eredmények is alátámasztanak. Ebben az esetben 9 VOTAT és 3 nem VOTAT, de elegendő információt szolgáltató minimális stratégia különíthető el egymástól. A korábbi kódolást alkalmazva VOTAT-stratégiák: +1+1+1; +1+1+2; +1+1+3; +1+2+1; +1+2+2; +1+2+3; +1+3+1; +1+3+2; +1+3+3; nem-VOTAT, de helyes stratégiák: +1+4; +4+2; +4+3. Továbbra is a leghatékonyabb stratégia a teljes mértékben elszigetelt változókezelést alkalmazó VOTAT-stratégia: +1+1+1 (5.20. ábra) volt. A leggyakrabban alkalmazott és az alacsonyabb komplexitású rendszerekben is a leghatékonyabbnak bizonyult VOTAT-stratégiát alkalmazó diákok közel 90%-a helyesen rajzolta fel a három bemeneti és két kimeneti változóval rendelkező problémák belső összefüggésrendszerét. A további nyolc VOTAT-stratégia közül csak hat került alkalmazásra, a +1+2+2 stratégia alkalmazása fordult még elő relatív gyakran (a diákok 10%-a alkalmazta ezt).



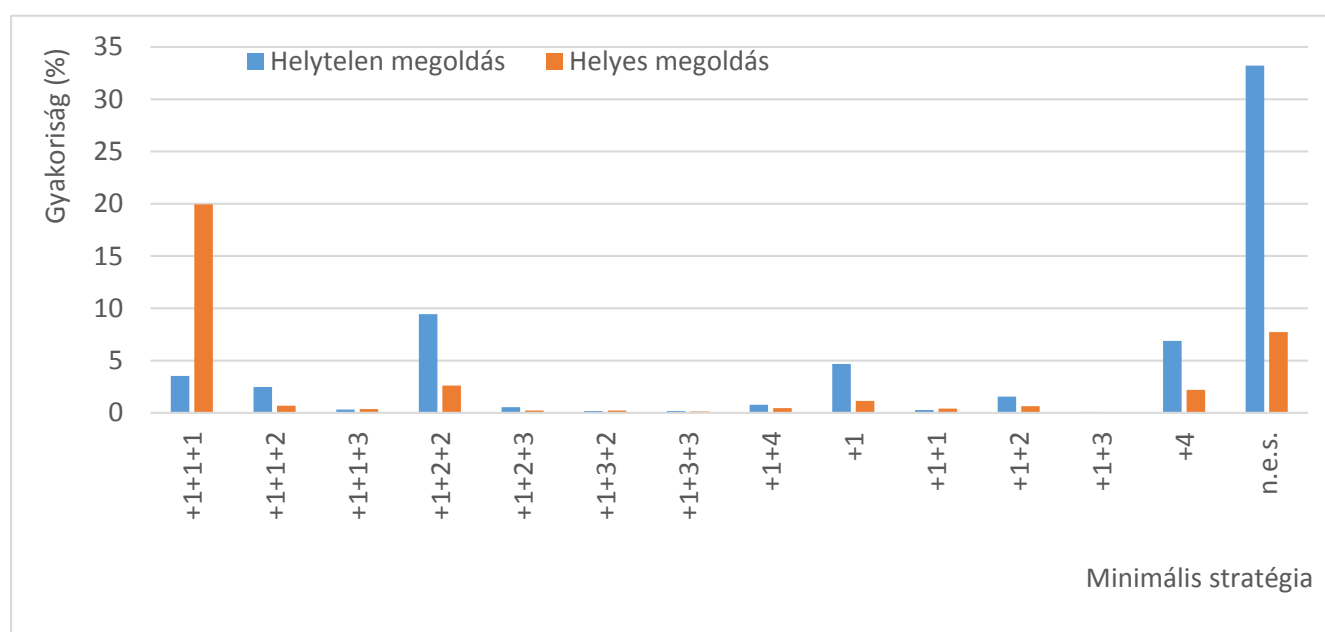
5.20. ábra

*A három bemeneti és két kimeneti változót tartalmazó problémák feltérképezése során alkalmazott helyes VOTAT és nem VOTAT minimális stratégiák a teljesítmény függvényében*

Annak ellenére, hogy helyes stratégiaalkalmazásról van szó, a diákoknak kevesebb mint harmada tudta helyesen felépíteni a modellt az így kinyert információk segítségével. Az összes többi VOTAT és helyes nem VOTAT-stratégia alkalmazása elhanyagolható mennyiségben és alacsony hatékonysággal fordult elő. A nem elegendő stratégiák között külön kiemeltük a +1, a +1+1, a +1+2, +1+3 és +4-es stratégiákat. Ezek két bemeneti változó esetén hatékony stratégiák voltak, de egy három bemeneti változóból álló rendszer feltérképezése esetén már nem elegendők. Közöttük a +1 és a +1+2-es stratégia gyakorisága volt kiemelkedő. Három bemeneti változó esetén még tovább csökkent a nem elegendő stratégiát alkalmazó, mégis intuitíven helyes problémamegoldónak bizonyuló tanulók aránya.

A három bemeneti és három kimeneti változót tartalmazó problémákon alapuló elemzések alátámasztották a korábbi eredményeket (5.21. ábra). A VOTAT-stratégiák közül a problémát

jól megoldók között továbbra is kiemelt helyet foglalt el a +1+1+1 minimális stratégia. Az ezt alkalmazók több mint 80%-ban helyesen oldották meg a problémát, azaz sem a helyes problémamegoldás, sem a helyes stratégia külön-külön nem ad teljes képet a diákok problémamegoldó gondolkodásának fejlettségi szintjéről. A +1+2+2 stratégia alkalmazása fordult még elő gyakrabban a VOTAT-stratégiák között, azonban ebben az esetben feltételezhető, hogy a diákok figyelmen kívül hagyták, hogy kumulált hatást tesztelnek, ezért kevesebb, mint harmaduk hozott helyes döntést a rendszer felépítését tekintve. A +1+2+2 minimális stratégiával dolgozó problémamegoldók jelentős része helytelenül rajzolta fel a rendszer szerkezetét mutató modellt, miközben a helyes modellépítéshez minden információ rendelkezésükre állt. A többi VOTAT és helyes, de nem VOTAT-stratégia alkalmazása elhanyagolható szerepet játszott. A nem elegendő stratégia alkalmazás már elenyésző esetben vezetett helyes megoldáshoz, azaz a változók és kapcsolatok számának növekedésével közel teljes mértékben visszaszorult a helyes találgatások aránya és minden egyes nem elegendő stratégia esetén (+1, +1+1, +1+2, +1+3, +4, n.e.s) gyakoribb volt a helytelen, mint a helyes válasz.



5.21. ábra

*A három bemeneti és három kimeneti változót tartalmazó problémák feltérképezése során alkalmazott helyes VOTAT és nem VOTAT alapstratégiák a teljesítmény függvényében*

A diákok által alkalmazott stratégiaelemzések rávilágítottak arra, hogy a teszt megbízhatósági mutatói jelentős mértékben növekednek, ha nemcsak a diákok konkrét válaszaikra alapozunk az elemzéseket, hanem elemezzük, hogy az általuk alkalmazott stratégia segítségével megoldható-e a probléma. Annak ellenére, hogy a MicroDYN rendszerekhez hasonló minimális komplexitású rendszerek kiismerése, feltérképezése számos problémamegoldó stratégiával megvalósítható, mégis a szakirodalomban közel egyedülként tárgyalt VOTAT-stratégiákhoz sorolható, bementi változók számától függő +1+1, illetve +1+1+1 minimális stratégiák alkalmazása bizonyult a leghatékonyabbnak. Az alacsony számú (két bementi és egy kimeneti) változóval rendelkező rendszerekben magas volt a helyes

találgatás aránya, ugyanakkor még azon egyszerű rendszerben is relatív magas arányban képviseltették magukat azon diákok, akik helyes stratégiát alkalmaztak a rendszer feltérképezése során, azaz sikerült kinyerniük az összes lényeges információt a rendszer működéséről, de azt már nem tudták értelmezni, felrajzolni, reprezentálni egy modellben. Már a két bementi és két kimeneti változóval rendelkező problémák esetén is jelentős mértékben visszaesett a helyes találgatás, az intuitív problémamegoldók aránya és megjelent az a tendencia, miszerint a számos elegendő információt szolgáltatató VOTAT és nem VOTAT-stratégia ellenére is egy speciális VOTAT-stratégia alkalmazása vezetett leginkább helyes megoldáshoz. A teljes mértékben elszigetelt változókezeléssel dolgozó stratégia alkalmazása sem vezetett minden esetben helyes megoldáshoz, ezért felmerült a kérdés, hogy feltérképezhető-e e stratégia alkalmazásának tudatossága és a tudatos stratégiahasználók is hasonló arányban oldják-e meg e problémákat. A többi, elméletileg jó stratégia alkalmazása esetén minden esetben magasabb volt a helytelen értelmezés, a helytelen válaszok aránya. A +1+1, illetve +1+1+1 stratégia mellett arányaiban leggyakrabban a +1+2, illetve a +1+2+2 stratégia fordult elő, ami bár helyes stratégia, mégis a diákok 20-30%-a tudta csak helyesen értelmezni és reprezentálni az így kinyert információkat (valószínű az eredmények téves interpretálása miatt).

A logfájlelemzések és a felállított minimális, alap és teljes stratégia modellek segítségével választ kereshettünk arra a korábban említett kutatási kérdésre is, hogy vajon az alkalmazott stratégia tudatossága milyen mértékben befolyásolta a diákok problémamegoldó teljesítményét. Ennek feltérképezése céljából összevetettük a diákok által alkalmazott alap és minimális stratégiákat. A tudatosság detektálása során az adott stratégia időben és egymásutániségében összetartozó lépéseit kerestük. Míg az alapstratégia tartalmazza az időbeniséget (kizárja a teljes stratégiában még előforduló ismétlődéseket), a kipróbálás sorrendjét, addig a minimális stratégia már időbeliség nélkül kezeli a kivitelezett lépéseket, viszont információt szolgáltat arról, hogy a folyamat során kipróbálta-e a diák az adott lépéskombinációt.

A tudatosság elemzése során az alkalmazott VOTAT-stratégiákat egyben kezeltük, illetve kiemeltük +1+1, és +1+1+1, illetve sajátdinamika esetén a sajátdinamikát vizsgáló (+A) lépéssel kiegészített, de azonos stratégiákat. Az eredmények értelmében (5.7. táblázat) jelentős különbség van a felfedezés során valamely VOTAT-stratégiát alkalmazók és az egymás utáni lépések alapján tudatos VOTAT-stratégiát alkalmazók száma és teljesítménye között.

A tudatos alkalmazók jóval nagyobb arányban reprezentálták a rendszerből kinyert információkat helyesen. A legkisebb komplexitású feladatok esetén a tudatos +1+1 stratégiahasználók között négyszer annyian hoztak helyes döntést, mint helytelen. Ez az arány két, illetve két és félszeres volt az időfaktort figyelmen kívül hagyó elemzésekben. A bonyolultabb rendszerek esetén is tapasztalható volt ez a tendencia, – hét - hét és félszer annyian hoztak helyes döntést a tudatos stratégiahasználók, mint a kevésbé tudatosak (öt és fél - hat ez az arány) – bár kisebb mértékben, miután a logfájlelemzések eredményei alapján a bonyolultabb rendszereknél kisebb arányban fordultak elő a VOTAT-stratégiát nem tudatosan alkalmazó diákok.

A sajátdinamikával rendelkező problémák feltérképezése kapcsán a VOTAT-stratégiát alkalmazó és helyes döntést hozó diákok mindegyike tudatos stratégiaalkalmazónak bizonyult. Kivétel nélkül a helytelen megoldást adó csoportból kerültek ki a kevésbé tudatos stratégiahasználók.

5.7. táblázat. A tudatos és kevésbé tudatos VOTAT-stratégia használók problémamegoldó sikeressége

Komplexitás/VOTAT-stratégia		Minimális stratégia			Alap stratégia			Eltérés		
		0	1	Össz.	0	1	Össz.	0	1	Össz.
2-1	Sum_VOTAT	410	957	1367	326	801	1127	84	156	240
	+1+1	191	498	689	68	259	327	123	239	362
2-2	Sum_VOTAT	594	708	1302	466	615	1081	128	93	221
	+1+1	212	461	673	76	307	383	136	154	290
3-2	Sum_VOTAT	334	487	821	277	459	736	57	28	85
	+1+1+1	65	380	445	47	353	400	18	27	45
3-3	Sum_VOTAT	364	528	892	305	502	807	59	26	85
	+1+1+1	77	435	512	67	418	485	10	17	27
Sajátdinamika										
3-2	Sum_VOTAT+A	143	31	174	124	31	155	19	0	19
	A+1+1+1/+1+A+1+1/+1+1+A+1/+1+1+1+A	77	30	107	66	28	94	11	2	13
3-3	Sum_VOTAT+A	83	43	126	75	43	118	8	0	8
	A+1+1+1	62	41	103	61	41	102	1	0	1

Összességében megállapítható, hogy a kevésbé tudatos stratégiahasználók között (5.7. táblázat eltérés oszlopa) arányaiban nagyobb mértékben fordultak elő a helytelen megoldást adó diákok, a tudatos VOTAT-stratégiát használók pedig nagyobb arányban értelmezték helyesen és képezték le jól a rendszerből kinyert információkat. Ennek ellenére a tudatos stratégiahasználók között továbbra is voltak olyan diákok, akik nem tudták helyesen reprezentálni a rendszerből kinyert információkat, azaz a tudatos feltérképezés sem jelentett egyértelműen helyes megoldást. Minél bonyolultabb volt a probléma mögött húzódó rendszer, annál inkább elkülönült egymástól a tudatos és kevésbé tudatos VOTAT-stratégia használók teljesítménye és egyre inkább erősödött az a tendencia, hogy a tudatos stratégiahasználók jól is reprezentálják a rendszerből kinyert információkat, azaz helyes döntéseket hoztak a probléma megoldása során.

Miután a VOTAT-stratégiák között kiemelkedő szerepűnek bizonyult az elszigetelt változókezelésen alapuló stratégia (+1+1, illetve +1+1+1), ezért a különböző típusú problémamegoldók profiljainak meghatározását, a különböző típusú problémamegoldók csoportosítását e stratégia alapján végeztük. A korábbi stratégiaelemzések eredményei szerint nem volt jelentős eltérés a 6-os és a 8-os diákok stratégiahasználatának sikerességében ( $t=-1,93$ ,  $p>0,05$ ). Feltételezésünk szerint nemcsak ezek a minimális mennyiségi, hanem minőségi eltérések is definiálhatóak a különböző típusú stratégiahasználók között és egymástól jól elkülöníthetőek a különböző típusú problémamegoldók.

Az elszigetelt változókezelés megvalósítását néző elemzésekbe nem vettük bele a sajátdinamikával rendelkező problémákat, miután azok nem térképezhetőek fel a +1+1 vagy +1+1+1 stratégiák alkalmazásával (3 db probléma), valamint nem vettük bele a legegyszerűbb, a korábbi elemzések alapján a helyes találgatás magas valószínűsége miatt kicsit máshogy viselkedő problémát sem. Ezért összesen hat probléma első fázisa logadatainak elemzésére került sor.



A problémamegoldók látens osztályaink meghatározását látens profil elemzéssel végeztünk. Az elemzésbe bevont kategorikus látens változókat a logfájlok alapján generált 0,1,2 kategóriákat tartalmazó változók adták. Kettőtől nyolcosztályos megoldásokat vizsgáltunk.

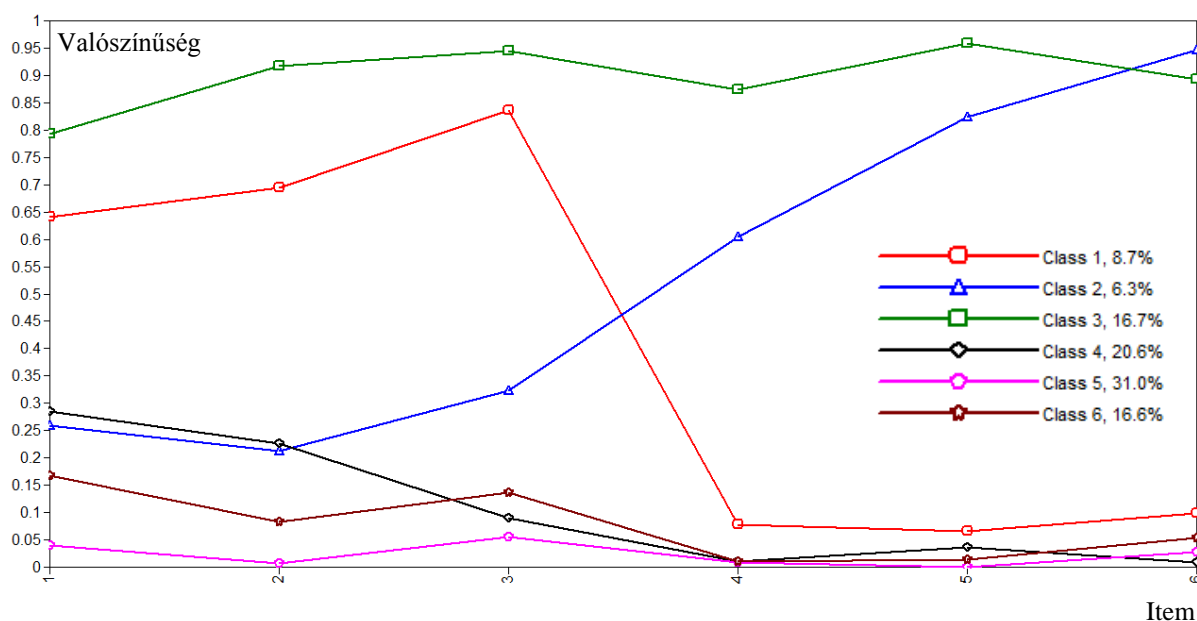
Az 5.8. táblázat mutatja a látens osztályok számától függően az információ alapú kritériumokat és az entrópiát. Mind az AIC, a BIC és az elemszámra kontrollált aBIC is folyamatos csökkenést mutatott az egyes osztályok hozzáadásával. A hatosztályos megoldás után azonban kiegyenlítődés mutatkozott. Az entrópiát tekintve a kettő és háromcsoportos megoldások érték el a maximum szintet. Az L-M-R Teszt (Lo-Mendell-Rubin valószínűségi hányados teszt) alapján azonban a hat egymástól elkülöníthető profilt tartalmazó modellt fogadtuk el.

5.8. táblázat. A látens profil elemzések illeszkedési mutatói a 6-tételes stratégiaelemzés esetén (6-8-os minta)

A látens osztályok száma	AIC	BIC	aBIC	Entrópia	L-M-R teszt	p
2	12043	12168	12089	0,910	2305	0,0001
3	11041	11232	11111	0,872	1016	0,0001
4	10902	11158	10996	0,829	162	0,0001
5	10827	11148	10944	0,832	100	0,0001
6	10761	11148	10903	0,824	90	0,0001
7	10748	11199	10914	0,802	39	0,3162
8	10746	11263	10936	0,806	27	0,4608

Megj.: AIC: Akaike Információs Kritérium, BIC: Bayesi Információs Kritérium, aBIC: korrigált Bayesi Információs Kritérium, L-M-R teszt: Lo-Mendell-Rubin valószínűségi hányados teszt

A hat, egymástól elkülöníthető problémamegoldó profil jellemzőit az 5.22. ábra mutatja. A diákok köze harmada (31%) a legelemibb rendszereket sem kezelő problémamegoldók csoportjába tartozott (az ábrán class 5 jelölés, 1. profil 1), az ide sorolt diákok 92%-os biztonsággal jellemezhetők ezzel a problémamegoldó profillal. Tőlük kicsit sikeresebbek (class 6, profil 2) a legegyszerűbb rendszereket alacsony fokon átlátó problémamegoldók (16%). A jelen modellben ide sorolt diákokról 88% biztonsággal jelenthető ki, hogy ebbe és nem egy másik profilú csoportba tartoznak. Őket az alkalmazott problémamegoldó stratégiák alapján az egyszerű problémákat kisebb sikerekkel átlátó problémamegoldók (21%) követték (profil 3). Az e három profillal jellemezhető problémamegoldók, a diákok közel 70%-a, alapvetően nagyon alacsonyszintű képességekkel rendelkeznek és a teszt megoldása közben sem tanultak.



5.22. ábra

*A problémamegoldó stratégiákra alapuló látens profil elemzés (Jelmagyarázatban lévő feliratok: Class 5: A legegyszerűbb rendszereket sem kezelő problémamegoldók; Class 6: A legegyszerűbb rendszereket alacsony fokon átlátó problémamegoldók; Class 4: Egyszerű problémákat kisebb sikerekkel átlátó problémamegoldók; Class 2: A gyorsan tanuló problémamegoldók; Class 1: Az egyszerű rendszereket jól átlátó, de a bonyolultabbnál alulteljesítő problémamegoldók; Class 3: Magas szintű stratégiahasználók, problémamegoldók)*

A profilelemzés alapján a 6-8. évfolyamosok kb. 6%-a nagyon hatékony tanulási képességgel rendelkezik, ugyanis, míg a teszt elején az egyszerűbb rendszerek áttekintésével, feltérképezésével is problémájuk volt, addig az utolsó, legbonyolultabb (sajátdinamikával még nem rendelkező) problémát már közel 95% valószínűséggel oldották meg (class 2, profil 4). E gyorsan tanuló diákok 85%-os biztonsággal sorolhatóak e profilba. A diákok közel 9%-a (class 1) az egyszerű rendszereket jól átlátta, de a bonyolultabbnál már az alulteljesítő problémamegoldók csoportjához tartozott (profil 5). Végül a 6-8. évfolyamosok közel 17%-a magas szintű problémamegoldó képességekkel (profil 6), jó problémamegoldó stratégiákkal rendelkezett, ők azok, akik hatékonyan térképezték fel mind az egyszerű, mind a legbonyolultabb rendszereket is. 95%-os biztonsággal állítható ezen diákokról, hogy ez a profil jellemző rájuk (5.9. táblázat) és nem másik csoportba sorolhatóak.

5.9. táblázat. Az adott profilhoz tartozás valószínűsége

Látens osztály	Látens osztály					
	class 1 (profil 5)	class 2 (profil 4)	class 3 (profil 6)	class 4 (profil 3)	class 5 (profil 1)	class 6 (profil 2)
1	0,779	0,013	0,032	0,067	0,002	0,106
2	0,018	0,846	0,092	0,010	0,008	0,026
3	0,011	0,040	0,949	0,000	0,000	0,000
4	0,035	0,001	0,000	0,827	0,093	0,044
5	0,001	0,001	0,000	0,071	0,920	0,006
6	0,047	0,012	0,000	0,056	0,006	0,878

Az évfolyamonkénti elemzések egyrészt alátámasztották a teljes mintán végzett látens profilelemzések eredményeit, másrészt megmutatták a változás tendenciáit (5.10. táblázat). Előzetes hipotézisünkkel ellentétben nőtt a legalacsonyabb szintű stratégiahasználók aránya, ami a minta összetételéből is adódhatott, miután évfolyamonkénti bontásban nem reprezentatív mintaválasztásról volt szó.

5.10. táblázat. A különböző profillal jellemezhető problémamegoldó csoportokhoz tartozás gyakorisága évfolyamonkénti bontásban

Profilok	Évfolyam (gyakoriság, %)			Sum
	6	7	8	
A legelemibb rendszereket sem kezelő problémamegoldók	28,5	32,5	31,8	31,0
A legegyszerűbb rendszereket alacsony fokon átlátó problémamegoldók	17,1	10,4	12,5	16,6
Egyszerű problémákat kisebb sikerekkel átlátó problémamegoldók	26,8	28,2	14,1	20,6
A gyorsan tanuló problémamegoldók	3,7	4,4	5,7	6,3
Az egyszerű rendszereket jól átlátó, de a bonyolultabbnál alulteljesítő problémamegoldók	11,3	8,0	12,4	8,7
Magas szintű stratégiahasználók, problémamegoldók	12,5	16,5	23,5	16,7

Jelentős mértékben csökkent a másik két átlag alatt teljesítő profillal rendelkező diákok aránya. Kis mértékben nőtt a gyorsan tanulók és az egyszerű rendszereket átlátó, de a bonyolultabbakkal kevésbé boldoguló stratégiahasználók aránya. Végül hatodikról nyolcadik évfolyamra jelentős mértékben, közel kétszeresére növekedett a magas szintű stratégiahasználattal rendelkező diákok előfordulása. Az ő eredményeik járulhattak hozzá főképp az évfolyamonkénti átlagos teljesítményben tapasztalt növekedéshez.

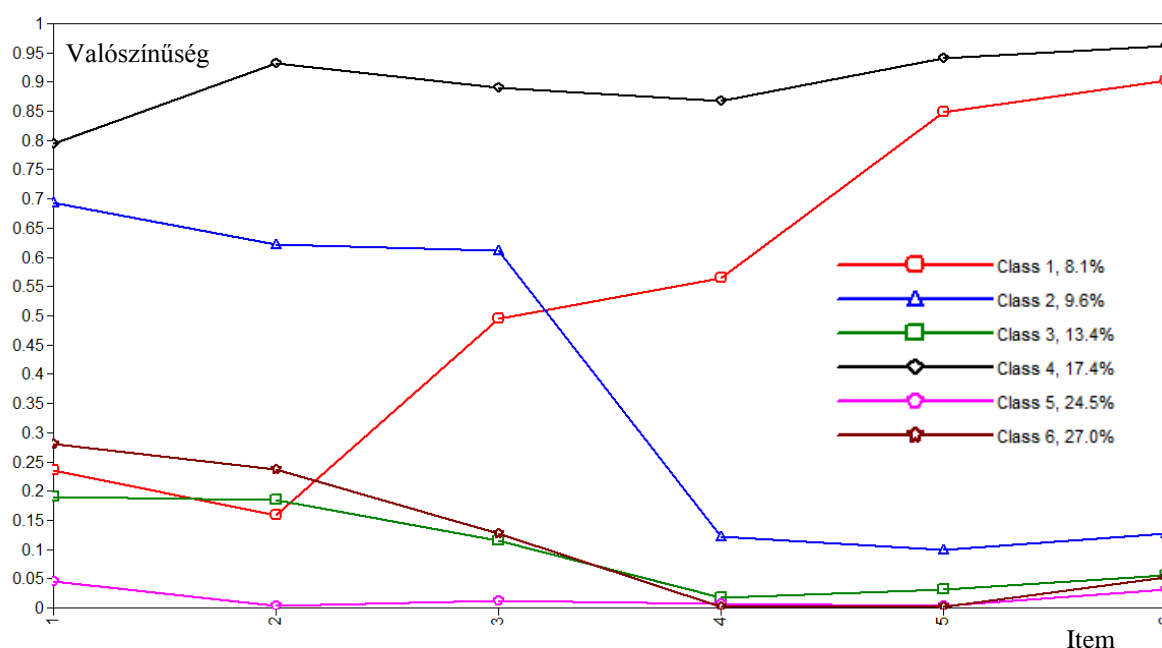
A fenti eredmények megerősítése céljából elvégeztük a látens profilelemzéseket a 2014-es adatfelvétel kizárólag hetedik évfolyamos diákokból álló, longitudinális kutatások keretein belül felvett mintáján. Az elemzések alapvetően alátámasztották a hat, egymástól elkülöníthető profilt (5.11. táblázat). Igaz, a hét csoportból álló modell jobban illeszkedett az adatokhoz, de valójában a legalacsonyabban teljesítő diákokat osztotta még két csoportra az elemzés (v.ö. 5.23. és 5.24. ábra).

A diákok több mint 60%-a továbbra is a legalacsonyabban teljesítő három csoporthoz tartozott. E három csoporton belüli elrendezésben tapasztalhatunk egyedül jelentősebb százalékbeli változást a két 2014-es adatfelvétel eredményeiben. A 25-13-27%-os eloszlás helyett 31-16-20%-os eloszlást, ami azonban a három csoport egymáshoz közeli jellemzőkkel leírható profilja miatt nem jelentős eltérés. A nagyobb mintán felvett adatok megerősítették a gyorsan tanuló diákok (6-8%), az egyszerű problémák jól, de a bonyolultabbakon alacsonyan teljesítők (8-10%) és a magasszintű felfedező és problémamegoldó stratégiákkal rendelkező diákok arányát (17-18%) a 7. évfolyamos mintában.

5.11. táblázat. A látens profil elemzések illeszkedési mutatói a 6-tételes stratégiaelemzés esetén (7-es longitudinális minta)

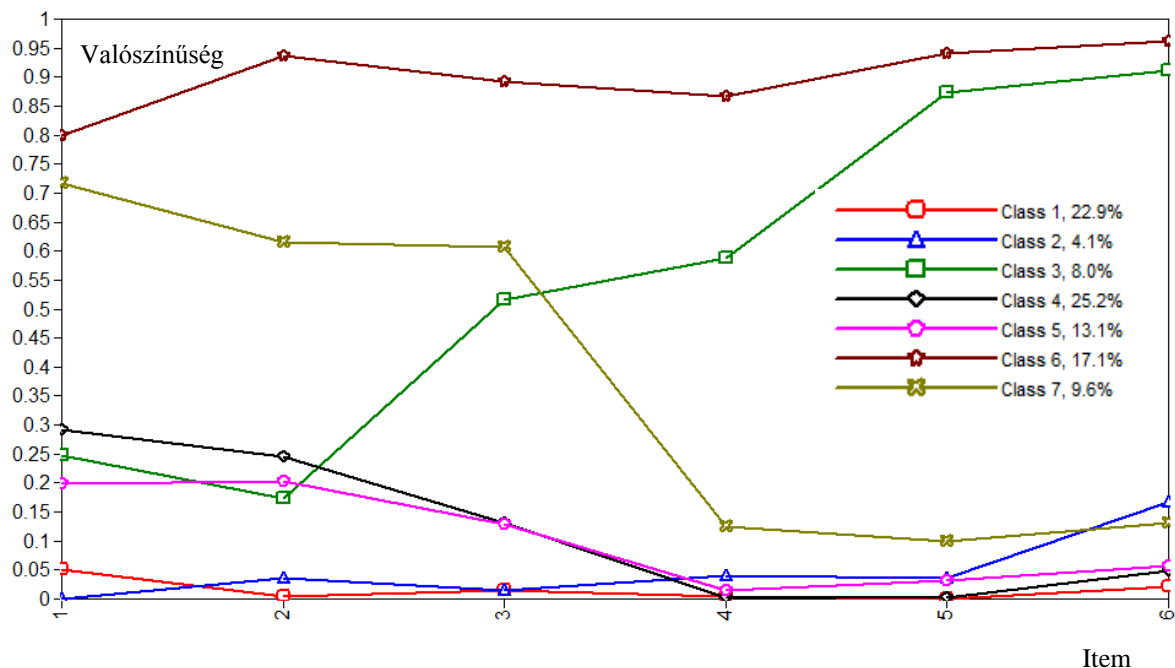
A látens osztályok száma	AIC	BIC	aBIC	Entrópia	L-M-R teszt	p
2	23093	23234	23155	0,893	4297	0,0001
3	21474	21689	21568	0,853	1628	0,0001
4	21254	21543	21381	0,790	243	0,0001
5	21111	21474	21270	0,791	167	0,0001
6	21042	21477	21233	0,771	94	0,0343
7	20996	21505	21219	0,774	71	0,0020
8	20992	21575	21248	0,788	13	0,1364

Megj.: AIC: Akaike Információs Kritérium, BIC: Bayesi Információs Kritérium, aBIC: korrigált Bayesi Információs Kritérium, L-M-R teszt: Lo-Mendell-Rubin valószínűségi hányados teszt



5.23. ábra

A problémamegoldó stratégiákra alapuló látens profil elemzés (Jelmagyarázatban lévő feliratok: Class 5: A legegyszerűbb rendszereket sem kezelő problémamegoldók; Class 3: A legegyszerűbb rendszereket alacsony fokon átlátó problémamegoldók; Class 6: Egyszerű problémákat kisebb sikerekkel átlátó problémamegoldók; Class 2: A gyorsan tanuló problémamegoldók; Class 1: Az egyszerű rendszereket jól átlátó, de a bonyolultabbnál alulteljesítő problémamegoldók; Class 3: Magas szintű stratégiahasználók, problémamegoldók)



5.24. ábra

*A problémamegoldó stratégiákra alapuló hét látens profilt definiáló elemzés csoportjai*

A fejlődési folyamatok jellegének pontosabb megértése és értelmezése végett lefuttattuk a látens profilelemzéseket az egyetemista korosztályban felvett logfájlokon is. A középiskola után erős szelekción átesett és az ország egyik vezető egyetemén első évfolyamot kezdő diákok problémamegoldó képességeik és az alkalmazott stratégiák alapján az L-M-R teszt eredménye szerint négy, különböző profillal jellemezhető csoportba sorolhatók (5.12. táblázat). A diákok 11%-a kisebb sikereket ért el az egyszerűbb problémák megoldása során, de a komplexebbekkel már nem boldogultak. A diákok egy másik 10%-a az egyszerűbb problémák megoldása során képes volt fejlődni, tanulni a teszt megoldása során, de a bonyolultabb, komplexebb rendszereket már nem látták át. A diákok 17%-a sorolható a gyorsan tanulók közé, ők azok, akik a teszt elején csak kisebb sikereket értek el az egyszerűbb problémák megoldása során, ugyanakkor a teszt végére megtanulták kezelni még a legbonyolultabb rendszereket is. Végül, a diákok 62%-a a legmagasabb szintű problémamegoldók és stratégiahasználók közé sorolható (5.25. ábra).

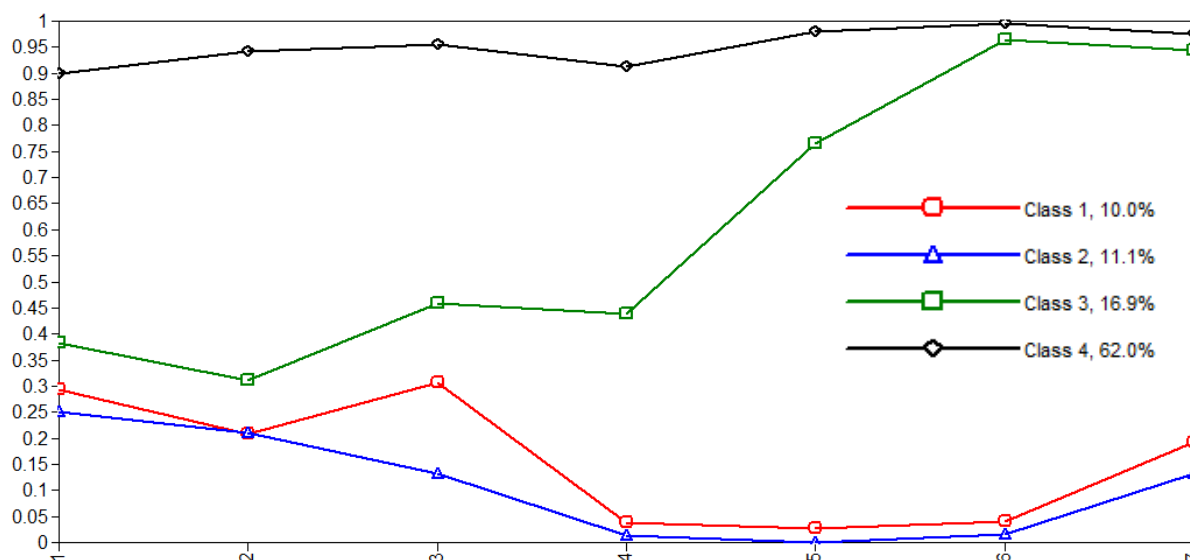
Ezek az eredmények megerősítették korábbi megállapításainkat, miszerint bár oktatási rendszerünk kevésbé fejleszti a diákok gondolkodási képességeit, beleértve problémamegoldó gondolkodásukat, de az iskolai szelekció jelentős mértékben magyarázható a diákok problémamegoldó képességeinek fejlettségi szintjével. Az egyetemista korosztályban mindössze 20% volt azon diákok aránya, akik gondolkodási képességük alapján a legalacsonyabban teljesítők közé sorolhatóak.

5.12. táblázat. A látens profil elemzések illeszkedési mutatói a diákok problémamegoldó stratégiahasználatának elemzése alapján (egyetemista minta)

A látens osztályok száma	AIC	BIC	aBIC	Entrópia	L-M-R teszt	p
2	10071	10220	10128	0,960	3344	0,0001
3	9611	9837	9697	0,938	486	0,0001
4	9337	9640	9453	0,888	300	0,0001
5	9287	9667	9432	0,896	79	0,8848
6	9267	9724	9441	0,893	49	0,8697
7	9253	9788	9457	0,894	43	1,0000

Megj.: AIC: Akaike Információs Kritérium, BIC: Bayesi Információs Kritérium, aBIC: korrigált Bayesi Információs Kritérium, L-M-R teszt: Lo-Mendell-Rubin valószínűségi hányados teszt

A logfájlokkal végzett látens profilelemzések lehetőséget biztosítottak arra, hogy kvalitatív szempontból jellemezzük a diákok által dinamikus problémamegoldó helyzetekben alkalmazott felfedező stratégiákat. Az eredmények alapján egy a legelemibb rendszereket sem kezelő, a legegyszerűbb rendszereket alacsony fokon átlátó, az egyszerű problémákat kisebb sikerekkel átlátó, a gyorsan tanuló, az egyszerű rendszereket jól átlátó, de a bonyolultabbnál alulteljesítő és a magas szintű stratégiahasználók, problémamegoldók hatosztályos rendszerrel írhatóak le, jellemezhetőek leginkább az adatok. Keresztmetszetileg a magasabb évfolyamra járó diákok csak kis mértékben mutattak szofisztikáltabb felfedező stratégiákat, nagyon alacsony mértékű volt a 6. és 8. évfolyam között megfigyelhető fejlődés.



5.25. ábra  
Egyetemisták problémamegoldó stratégiáinak látens profilelemzése

Eredetileg, a szakirodalom alapján egy négyosztályos eredményt feltételeztünk. Három osztályt a diákok teszten nyújtott teljesítménye alapján: alacsony szintű, a közepes és a magas szintű problémamegoldó, azaz minél magasabb szintű felfedező stratégiákkal rendelkezik, annál jobban teljesít a teszten; valamint az alapvetően hasonló struktúrájú problémák miatt feltételeztük, hogy lesz egy, a gyorsan tanuló diákokat tömörítő osztály is.

A gyenge, közepes és magas szintű hármas osztályozás a szakértősség mennyiségi különbségeivel is megfogható, amire korábbi kutatások már felhívták a figyelmet (*Sonnleitner és mtsai*, 2012; *Greiff, Wüstenberg és Avvisati*, 2015; *OECD*, 2014). Ugyanakkor a tipikus közepes szintű problémamegoldó profil meglétét nem támasztották alá elemzéseink. A mennyiségi jellemzőkkel is leírható csoportokhoz tartozó diákok vagy a kiemelkedően jó (átlag\_evf7\_teszt=4,62; átlag\_evf7\_stratégia=5,8), vagy összességében az alacsony szintű problémamegoldókhoz tartoztak (átlag\_evf7\_teszt=1,32; átlag\_evf7\_stratégia=1,08).

A látens profilelemzés alapú megközelítés igazi előnye, hogy rávilágított arra, megerősítette azon hipotézisünket, hogy a fejlődés nemcsak mennyiségi, hanem minőségi változással is leírható. A stratégiahasználat szakértőssége kapcsán nem elegendő egy kizárólag mennyiségi elvű elemzés, sőt jelen esetben téves következtetések levonását eredményezné. Például a gyorsan tanulók csoportja a teszt elején úgy viselkedett, mint a legalacsonyabb szintű problémamegoldók, majd a problémákkal dolgozva egy jelentős tanulási fázison, minőségi változáson estek át és a teszt második felére már a magas szintű problémamegoldóknál tapasztalt teljesítményt mutatták. Egy kizárólag mennyiségi elvű elemzés a kezdeti alacsony teljesítmény miatt nem mutatná ki a valóságban elért és mutatott szakértőségi szintet, miután átlagban a szakértő problémamegoldók eredményei alatt teljesítettek. Sőt összességben az alacsony és a magas szintű problémamegoldók teljesítménye között definiálható átlagos stratégiahasználatuk (átlag\_evf7\_teszt\_gyorsan\_tanuló=3,09; átlag\_evf7\_stratégia\_gyorsan\_tanuló=5,17) miatt őket sorolná a rendszer a közepes teljesítményű problémamegoldók közé, akik viszont profilelemzéseink alapján nem alkotnak külön osztályt. A másik, átlagos teljesítménye alapján ide sorolt osztály (a diákok 9%-a) a kezdetben magasan teljesítők, majd a komplexebb rendszereken alulteljesítők csoportja lenne (átlag\_evf7\_teszt=2,02; átlag\_evf7\_stratégia=5,00). Az ő átlagos teljesítményük szintén közepes szintű, azonban jelentős minőségi eltérés van a két csoporthoz tartozó diákok között.

A gyors tanulók feltételezhetően sok, igény szerint, speciális helyzetekben könnyen adaptálható kognitív sémával rendelkeznek (*Markman*, 1999), ugyanakkor az adaptációhoz időre van szükségük. A szakértő problémamegoldók velük szemben sokkal kidolgozottabb, explicitebb sémákat érhetnek el, amelyekre alapozva már a kezdetektől fogva képesek a legoptimálisabb felfedező stratégia alkalmazására. A kezdetben magasan, majd bonyolultabb rendszerek esetén alacsonyan teljesítőknél feltételezhetően a kognitív túlterhelés jelent meg.

A diákok közel 40%-át a profilelemzés további két, korábban nem feltételezett csoportba sorolta. Az első csoport tagjai a teszt elején, a legegyszerűbb problémák kapcsán kisfokú tanulásról tettek bizonyosságot, majd a bonyolultabb rendszereknél egyértelműen a legalacsonyabban teljesítők csoportjába kerültek át. A diákok 21%-a már alacsony-közepes szinten teljesített a teszt elején, majd semmiféle fejlődést nem mutatva a problémamegoldó helyzetek komplexitásának növekedésével párhuzamosan fokozatosan a legalacsonyabban teljesítők táborához csatlakoztak.

Az évfolyamok közötti fejlődés az elvártnál, a feltételezettnél kisebb mértékű volt. Az egyedüli jelentős változást a közel kétszeresére növekedett magas szintű stratégiahasználattal rendelkező diákok előfordulása jelentette, ugyanakkor még 8. évfolyamon is a diákok 45%-a a két legalacsonyabb képességszintű csoportba tartozott. Ők azok, akik gyakorlatilag semmilyen szinten sem képesek az elszigetelt változókezelés alkalmazására.

Az egyetemisták között történt adatfelvétel eredményeire építő elemzés alapvetően alátámasztotta a korábbi megállapításainkat, bár a látens profilelemzések – feltételezhetően az erős szelekció miatt – már csak négy csoportba osztotta a diákokat: magasan teljesítők; gyorsan tanulók; kezdetben átlag alatt teljesítők, de az egyszerű rendszereken tanulni képes és végül a kezdetben átlag alatt, alacsonyabban teljesítők, majd a problémák bonyolultsági fokával

párhuzamosan egyre alacsonyabban teljesítő diákok. Feltételezésünk szerint egy azonos korosztályú reprezentatív mintán kialakultak volna a legalacsonyabban teljesítő osztályok is.

A diákok stratégiahasználatát és CPS teszten nyújtott teljesítményét számos képesség, egyéni jellemző meghatározza. A disszertációban korábban bemutatott, valamint a szakirodalom által is alátámasztott ilyen képességek például az általános gondolkodási képességek és a fluid intelligencia (pl.: *Greiff, Wüstenberg, Molnár, Fischer, Funke és Csapó*, 2013; *Wüstenberg és mtsai*, 2012). Ugyanakkor a látens profilelemzések rámutattak arra, hogy a különböző tanulási típusok, stílusok (*Jonassen és Grabowski*, 1993) területe is releváns kutatási kérdésként merülhet fel a kapcsán, hogy a diákok milyen felfedező stratégiákat használnak problémamegoldó helyzetekben és hogyan tanulnak a teszt megoldása során.

További logfájl-elemzéseket végeztünk a célból, hogy választ kapjunk arra a kutatási kérdésre, hogy a kivitelezett manipulációk száma és a probléma megoldásával töltött idő hogyan függ össze az alkalmazott stratégia helyességével és a teszten, problémákon mutatott teljesítményekkel. Az eredmények értelmében a teljesítmény, azaz a diákok válaszainak helyessége és a probléma feltérképezését célzó manipulációk száma és a probléma feltérképezésével töltött idő kevésbé függött össze. A logfájl-elemzések nélkül azt mondhatnánk, hogy a diákok teljesítményét, problémamegoldó hatékonyságukat kevésbé határozta meg a probléma kiismerésével eltöltött idő és a problémával való interakciók száma. A logfájl- és stratégiaelemzések azonban rávilágítottak arra, hogy mind a problémával való interakciók mennyisége, mint a probléma feltérképezésére szánt idő és a problémamegoldás során alkalmazott stratégia helyessége között közepes erősségű szignifikáns kapcsolat volt, azaz mégis fontos tényezők a problémamegoldás folyamatában. Minél bonyolultabb, komplexebb rendszer feltérképezéséről volt szó, annál erősebb volt ez a kapcsolat (5.13. táblázat). Míg a tesztben szereplő legegyszerűbb rendszerek esetén  $r=0,3-0,4$ , addig a három bemeneti változót tartalmazó problémák kapcsán már  $r=0,6$  feletti a probléma feltérképezésére szánt manipulációk mennyisége,  $r=0,5$  körüli a feltérképezésre szánt idő és a helyes stratégia alkalmazása közötti kapcsolat erőssége (mindegyik korrelációs együttható  $p<0,01$  szinten szignifikáns).

5.13. táblázat. A kivitelezett manipulációk száma és a probléma megoldásával töltött idő összefüggése az alkalmazott stratégia helyességével és a teszt egyes problémáin mutatott teljesítménnyel

Probléma száma	r (teljesítmény/ manipulációk száma)	r (teljesítmény/ idő)	r (elméletileg helyes stratégia/ manipulációk száma)	r (elméletileg helyes stratégia/ idő)
1.	n.s.	n.s.	0,453**	0,352**
2.	n.s.	-0,068**	0,477**	0,308**
3.	0,072**	0,073*	0,471**	0,326**
4.	0,047*	n.s.	0,483**	0,320**
5.	0,136*	0,098*	0,541**	0,380**
6.	n.s.	n.s.	0,376**	0,350**
7.	0,137**	0,097**	0,610**	0,431**
8.	0,149**	0,088**	0,555*	0,377**
9.	0,098**	0,124**	0,645**	0,446**
10.	0,183**	0,134**	0,667**	0,484**

Megj. A problémák komplexitását l. 6.3. táblázat; \*,  $p<0,05$ , \*\*  $p<0,01$  szinten szignifikáns.

A profilelemzések alapján kialakított csoportokba tartozó diákok átlagosan különböző mennyiségű időt töltöttek a problémák feltérképezésével. A problémák szerkezetének



megismerésével töltött idő profilonkénti bontásban alapvetően nem változott a 6–8 évfolyam vonatkozásában. A legalacsonyabb képességszintű diákok kétharmad annyi időt szántak a problémák kiismerésére (evf7: 258,34 másodperc), mint a legmagasabb képességszintűek (evf7: 364,67 másodperc). Az ANOVA elemzés eredménye szerint a legalacsonyabb képességszintűeknél szignifikánsan több időt, ugyanakkor egymás között azonos mennyiségű időt töltöttek a gyorsan tanulók (evf7: 346,52 másodperc), a szakértő problémamegoldók, az egyszerű problémákon kisfokú tanulást mutató, de a komplexeken kudarcot valló diákok (evf7: 331,10 másodperc) és az egyszerű problémákon alacsony-közepes szinten, a bonyolultabb rendszereken alulteljesítő diákok (evf7: 350,02 másodperc) is. A problémák feltérképezésére a legtöbb időt ( $p < 0,05$ ) az egyszerű problémákon jól, de a komplexeken alulteljesítő diákok (evf7: 450,00 másodperc) szánták.

A diákok által teszten belül alkalmazott manipulációk mennyisége közötti kapcsolat egyre erősödött (5.14. táblázat). Míg a teszt első két-három problémája feltérképezése kapcsán kevésbé volt azonos, addig a negyedik problémától kezdve jellemző volt, hogy aki több interakciót alkalmazott a rendszer megismerése során, az a következő problémánál is ezt tette, míg aki kevesebbet, az később sem járt el másképpen. Ez is alátámasztja korábbi eredményeinket, miszerint definiálhatók olyan típusú problémamegoldók, akiknek viselkedése alapvetően nem változik a teszt megoldása közben.

A disszertáció e fejezetében bemutatott elemzések egyértelműen rávilágítottak arra, hogy a számítógép-alapú tesztelés azon tulajdonsága, hogy logolhatóvá és elemezhetővé válnak a kontextuális adatok (pl.: kattintás, idő), olyan kutatási kérdések megválaszolását teszik lehetővé, amelyekre néhány évvel ezelőtt a hagyományos technikák alkalmazásával még nem tudtunk volna válaszolni. A logfájl-elemzések hozzájárultak a diákok által alkalmazott stratégiák pontosabb feltérképezéséhez, sőt az alkalmazott stratégiákban keresett mintázatok segítségével számszerűsíthetővé és definiálhatóvá váltak – a találgatás faktorának teljes kizárása mellett, kizárólag a kivitelezett szcenáriók elemzése alapján – a különböző típusú problémamegoldók.

5.14. táblázat. A teszt megoldása során az egymás utáni problémák feltérképezése során alkalmazott manipulációk száma közötti összefüggések

Kipróbált lehetőségek száma (alkalmazás gombra való kattintás száma)	Kipróbált lehetőségek száma (alkalmazás gombra való kattintás száma)									
	r	2. probl.	3. probl.	4. probl.	5. probl.	6. probl.	7. probl.	8. probl.	9. probl.	10. probl.
1. pr.		.384**	.362**	.372**	.356**	.314**	.271**	.277**	.259**	.265**
2. pr.	1		.485**	.456**	.395**	.266**	.341**	.319**	.368**	.349**
3. pr.			1	.570**	.504**	.480**	.423**	.449**	.416**	.395**
4. pr.				1	.569**	.501**	.455**	.487**	.420**	.402**
5. pr.					1	.654**	.553**	.573**	.506**	.489**
6. pr.						1	.497**	.612**	.386**	.367**
7. pr.							1	.584**	.581**	.567**
8. pr.								1	.615**	.600**
9. pr.									1	.705**
10. pr.										1

A teszt megbízhatósági mutatói jelentős mértékben növekedtek, ha nemcsak a diákok konkrét válasza, hanem az általuk alkalmazott és logolt stratégiákra is alapoztuk az elemzéseket. Annak ellenére, hogy a minimális komplexitású rendszerek feltérképezése számos

stratégiával megvalósítható, mégis a változók szigorú elszigetelésére alapozó VOTAT-stratégia alkalmazása bizonyult a leghatékonyabbnak. A többi, elméletileg jó stratégia alkalmazása esetén magasabb volt a helytelen értelmezés, a helytelen válaszok aránya. A diákok egy jelentős hányada elegendő információt szolgáltató stratégiát alkalmazott a rendszerek feltérképezése során, de azt már nem tudták értelmezni és felrajzolni a kért modellben.

A tudatos stratégiahasználók jóval nagyobb arányban reprezentálták a rendszerből kinyert információkat helyesen. A legkisebb komplexitású feladatok esetén négyszer annyian hoztak helyes döntést, mint helytelen. A bonyolultabb rendszerek esetén is tapasztalható volt ez a tendencia, bár kisebb mértékben, miután a logfájlelemzések eredményei alapján a bonyolultabb rendszereknél kisebb arányban fordultak elő a VOTAT-stratégiát nem tudatosan alkalmazó diákok. A sajátdinamikával rendelkező problémák feltérképezése kapcsán a VOTAT-stratégiát alkalmazó és helyes döntést hozó diákok mindegyike tudatos stratégiaalkalmazónak bizonyult. Minél komplexebb, több változót tartalmazott a feltérképezendő és megoldandó probléma mögött húzódó rendszer, annál inkább elkülönült egymástól a tudatos és kevésbé tudatos VOTAT-stratégia használók teljesítménye és egyre inkább erősödött az a tendencia, hogy a tudatos stratégiahasználók jól is reprezentálják a rendszerből kinyert információkat, azaz helyes döntéseket hoztak a probléma megoldása során.

A látens profilelemzések alapján egyértelműen hat különböző tulajdonságokkal jellemezhető csoport különíthető el egymástól: a legelemibb rendszereket sem kezelő problémamegoldók, ide sorolhatóak a 6-8. évfolyamos diákok harmada. Náluk kicsit sikeresebbek a legegyszerűbb rendszereket alacsony fokon átlátó és kisebb fokú tanulást mutató problémamegoldók csoportja (16%), akik azonban a bonyolultabb rendszereknél már a legalacsonyabban teljesítők eredményeit mutatják. Hozzájuk tulajdonságukban egyrészt közel állnak, de mégis külön profillal jellemezhetőek az egyszerű problémákat kisebb sikerekkel átlátó, de a bonyolultabb rendszerekkel nem boldoguló problémamegoldók (21%), ők ugyanis nem mutattak fejlődést, tanulást a teszt problémái megoldása közben. Azon diákok, akik e tulajdonságokkal leírható csoportok valamelyikébe sorolhatóak, alapvetően nagyon alacsony szintű problémamegoldó, feltérképező, explorációs képességekkel rendelkeznek (a diákok 70%-a) és tanulási képességük is alacsony szintű. Ezzel szemben a profilelemzés alapján a 6-8. évfolyamosok kb. 6-8%-a nagyon hatékony tanulási képességekkel rendelkezik. Ők azok, akiknek a teszt elején az egyszerűbb rendszerek áttekintésével, feltérképezésével is problémájuk volt, ugyanakkor a teszt megoldása közben megtanulták annak használatát, a minimálisan komplex rendszerek hatékony feltérképezésének módszerét. A teszt végére a legbonyolultabb (sajátdinamikával még nem rendelkező) problémákat is már közel 95% valószínűséggel oldották meg. Az egyszerű problémák megoldás kapcsán náluk jobban teljesített a diákok azon 8-10%-a, akik az egyszerű rendszereket jól átlátták, de a bonyolultabbnál már az alulteljesítő problémamegoldók csoportjához tartoztak. Végül a 6-8. évfolyamosok közel 17%-át sorolhattuk a magas szintű problémamegoldó képességekkel, jó problémamegoldó stratégiákkal rendelkező diákok közé, ők azok, akik hatékonyan térképezték fel mind az egyszerű, mind a legbonyolultabb rendszereket is. Az egyetemista korosztályban, az érettségit követő szelekció hatására arányuk meghaladja a 60%-ot, de a gyorsan tanuló diákok ráta is jelentősen, duplájára emelkedett a vizsgált mintán. A problémák feltérképezésével töltött idő és interakciók száma bár kis mértékben függött össze a teljes teszten nyújtott teljesítményükkel, ugyanakkor az alkalmazott stratégiákkal már közepes-erős szintű kapcsolatot jelzett. A kapcsolat erőssége változott a problémák komplexitásának

függvényében. Az egyszerű problémák esetén a korábban tapasztalt nagyfokú találgatás is hozzájárult a gyengébb kapcsolat meglétéhez, míg a bonyolultabb problémáknál egyértelműen kimutatható volt, hogy aki több időt töltött a rendszer feltérképezésével, több scenáriót alkalmazott, teljesítménye is magasabb volt. A látens profilelemzések összességében megerősítették azt az elméletet, miszerint a képességek fejlődése nem csak mennyiségi, hanem minőségi változással is jellemezhető és leírható. Ennek szükségessége elkerülhetetlen, miután a kizárólagos mennyiségi elemzések jelen esetben is téves következtetések levonását vonták volna maguk után.

## 6. ÖSSZEFOGLALÁS ÉS JÖVŐBELI FELADATOK

A disszertáció öt fejezetében összefoglaltuk a technológia oktatási integrációjának lehetőségeit és fontosságát; a pedagógiai mérés-értékelés technikáinak fejlődési tendenciáit, külön hangsúlyt fektetve a technológiaalapú tesztelésre való átállás előnyeire, hátrányaira, kihívásaira, elterjesztésének feltételrendszerére és a vonatkozó nemzetközi és hazai tendenciákra. Külön fejezetben tárgyaltuk a számítógép-alapú tesztelés bevezetésének nemzetközi szinten legkevésbé kutatott és leginkább kérdéses életkori intervallumban, kisiskolás korban történő alkalmazhatóságára vonatkozó kutatási eredményeket. Végül a problémamegoldó képesség mérési és értékelési lehetőségeinek változásán keresztül szemléltettük a számítógép-alapú tesztelésben rejlő mai lehetőségeket, miközben tág életkori intervallumban ismertettük a diákok egy 21. században kulcsfontosságú gondolkodási képességének, a problémamegoldó képességének fejlődését, a mennyiségi és minőségi változás fő tendenciáit és lehetséges okait. Az elemzések során nemcsak a klasszikus, hanem a valószínűségi tesztelmélet és a neveléstudományi kutatásokban jelenleg leginnovatívabb, új lehetőségeket adó strukturális egyenletek eszközrendszerét is alkalmaztuk. Az elemzett adatbázisok csak egy része vonatkozott a diákok konkrét válaszainak elemzésére, a dolgozatban nagy hangsúlyt fektettünk, a nemzetközi szinten is most megjelenő és ezért kiemelt fontossággal bíró logfájl-elemzésekre és a legmodernebb elemzési módszerek logadatok értelmezésében való felhasználására. A számítógép-alapú adatfelvétel és a modern elemzési eljárások ötvözése lehetőséget teremtett a vizsgált jelenségek alaposabb megértéséhez.

A disszertáció első fejezete részletesen kitért az információs és kommunikációs technológiák oktatásra gyakorolt hatására, változtatást indukáló erejére. A technológia oktatási integrációjának különböző lehetőségei és azok értékelése kapcsán kiemelt fontosságúnak tartjuk, hogy az érintett eszközök, azok innovatív felhasználását lehetővé tevő új módszertani repertoár, illetve mérés-értékelési kultúra együttes alkalmazása valósuljon meg. Azok egymagukban nem oldják meg a problémákat, illetve csak abban az esetben fejlesztik iskolai oktatásunk minőségét, ha az oktatás valós problémáinak megoldásában alkalmazzuk azokat. A technológia tanításban, tanulásban történő használata során alapvető fontosságú, hogy ne az eszközök határozzák meg a változtatások irányát, az a változtatások katalizátora legyen. Azon országokban, ahol a társadalom IKT fejlődési indexe magas, elmondható, hogy egyrészt a technológia az oktatás-tanulás szerves részévé vált, másrészt kivétel nélkül az OECD PISA mérések élvonalában helyezkednek el vagy a jelentős fejlődést elért országok között vannak.

A disszertáció második fejezete számos perspektívából ismertette a pedagógiai mérés-értékelés technikáinak fejlődési tendenciáit, miközben kitért a technológiaalapú mérésekben rejlő többlet lehetőségekre, előnyökre, hátrányokra és kihívásokra. A szakirodalmi elemzés egyik legfontosabb konklúziója, hogy a megfelelően alkalmazott modern technológia a tanulók tudásának rendszeres, objektív visszacsatoláson alapuló felmérésén, esetleges hiányosságaik időben történő diagnosztizálásán keresztül jelentős mértékben fokozhatja a mérés-értékelési, illetve tanítási-tanulási folyamat hatékonyságát. A technológiaalapú tesztelés alkalmazásával kitérítendő, szélesíthető a tesztelésbe bevont területek (pl.: IKT műveltség, digitális szövegek olvasása) és személyek köre (olvasni nem tudó diákok, sérült diákok). A valószínűségi tesztelmélet eszközrendszerén alapuló adaptív teszteléssel, az optimális kihívást biztosító feladatok személyre szabott kiköszvetítésével még pontosabbá, még motiválóbbá tehető a

diákok számára a tesztelés folyamata. Összességében megállapítható, hogy a technológia használata két különböző célból jelent meg a mérés-értékelés területén. A kutatások egyik felében azért történt meg a technológiai integráció, mert a modern technológiai eszközök nélkül nem lehetett volna felvenni az újonnan fejlesztett tesztek. A kutatások másik fele a technológia szélesebb körben való alkalmazására, a nagymintás mérésekbe történő bevezetésére fókuszált, ahol első feladat a papíralapú tesztelésről a számítógép-alapú tesztelésre való áttérés volt. A nemzetközi prominens mérés-értékelési projektekben (OECD PISA, IEA TIMSS) nyújtott teljesítményeink, kiemelten a számítógép-alapú teszteken mutatott eredményeink arra utalnak, hogy elodázhatatlan a számítógép-alapú mérés-értékelési kultúra minél szélesebb körű elterjesztése, ha továbbra is jelen szeretnénk lenni az európai mezőnyben. Az egyöntetű átállás akadályaként fogalmaztam meg az iskolai infrastruktúra sokszínűségét, a pedagógusok egy részének attitűdjét, és azt, hogy világszinten kevés olyan mérés-értékelési platform érhető el, melyek nemcsak a feleletválasztós itemek, hanem innovatív, 21. századi feladatok kezelésére, esetlegesen azonnali kiértékelésére is képesek. Hazánkban e feltételek részben adóttak: az iskolák infrastrukturális ellátottsága lehetővé teszi a kis tétellel bíró online diagnosztikus mérések bevezetését; a pedagógusok döntő többsége még a nagy tétellel bíró tesztek kapcsán is támogatja a számítógép-alapú tesztelésre való átállást; illetve rendelkezünk egy világszínvonalú, nemcsak hagyományos, hanem innovatív feladatok kezelésére is alkalmas és azonnali visszacsatolást biztosító online mérés-értékelési rendszerrel, az eDia-val. Az eDia rendszer a három fő műveltségi terület kapcsán: matematika, természettudományok és olvasás, közel 20.000 feladatot tartalmaz.

A harmadik fejezetben ismertetett empirikus kutatások célja annak feltérképezése volt, hogy a számítógép-alapú tesztelés alkalmazható-e kisiskolás korban és az eredmények hogyan viszonyulnak a korábbi hagyományos technikákkal történő adatfelvételek eredményeihez. Megfogalmazhatóak-e olyan megkorlátozások, olyan korlátozások, melyek kerülése javasolt a kisiskolás diákok számára kidolgozott tesztfeladatokban, miután azok teljesítménybefolyásoló erővel bírhatnak. A disszertáció e részében ismertetett géphasználattal és az iskolakészültég online vizsgálatával kapcsolatos kutatási eredmények a mintában lévő diákok fiatal életkora miatt nemzetközi szinten is hiánypótlónak tekinthetők. A diákok egér- és billentyűzethasználati képességét vizsgáló kutatások eredményei rávilágítottak arra, hogy bár az egy vagy több nagy elemre való kattintást, vonszolást, egy-két billentyű leütését, egy szó beírását kívánó feladatok nem jelentenek problémát a kisiskolás diákoknak, esetükben kerülendő azon típusú feladatok alkalmazása, ahol a válasz megadása során sok kicsi elemmel kell manipulálni, vagy egy-egy hosszabb szó, esetleg mondat, vagy billentyűzetkombináció segítségével elérhető, általában ritka karakterek begépelését kell megvalósítani. A teszt megoldása során jelentős mértékű tanulást detektáltunk, ami arra utal, hogy gyorsan és hatékonyan fejleszthető e képességük. Az iskolakészültég tesztek számítógépesítésére irányuló kutatások eredményei egyértelműen jelzik, hogy a technológiaalapú tesztelés nem csupán kivitelezhető ilyen fiatal korban, hanem sok tekintetben még alkalmasabb is, mint a személyes, szemtől szembeni adatfelvétel vagy papíralapú tesztelés. Az eredmények általánosíthatóságának korlátja, hogy a kutatásba bevont eredeti DIFER tesztek nem az első osztályt már megkezdett kisiskolás diákok, hanem az óvodából az iskolába átlépő gyerekek mérésére tervezték, ezért az elemzések során gyakran plafoneffektusba ütköztünk. Ennek ellenére e kutatási eredmények jó alapot biztosítottak egy új generációs, számítógép-alapú iskolakészültég tesztbateriá kidolgozásához, melyek munkálatai már folyamatban vannak.

Az egymásra épülő negyedik és ötödik fejezet egy konkrét gondolkodási képesség, a 21. században kiemelt szerepet kapó problémamegoldó képesség mérés-értékelési lehetőségeinek változásán keresztül mutatja be a számítógép-alapú tesztelésben rejlő mai lehetőségeket. Tág életkori intervallumban mind fejlődési, mind strukturális perspektívából elemeztük és hasonlítottuk össze a tudás elsajátításában és alkalmazásában kiemelt szerepet játszó két gondolkodási képesség, az induktív gondolkodás és a problémamegoldó képesség fejlődését, illetve a teszt közvetítő eszközének és egyéb (pl.: intelligencia, anyai iskolai végzettsége) háttérváltozóknak eredményekre, a vizsgált konstruktumokra gyakorolt hatását, prediktív erejét. A papíralapú tesztekkel végzett kutatások szintetizálásába közel 24000 diák válaszait vontuk be, míg a számítógéppel végzett, harmadik generációs tesztekkel kivitelezett kutatásokba több mint 7000 diák vett részt. A vizsgált gondolkodási képességek fejlődése a többi képesség fejlődési menetéhez hasonlóan lassú és jól jellemezhető egy négyparaméteres logisztikus görbével. A legintenzívebb fejlődési szakasz 6-8. évfolyamon tapasztalható. Extrapolálva a fejlődési folyamatokat, a 3. évfolyam előtt és a 11. évfolyam után is, bár lassuló ütemben, de folytatódik mindegyik vizsgált képesség fejlődése. A problémamegoldó gondolkodás fejlettségi szintjét előrejelző változók azonosítására irányuló elemzések rávilágítottak arra, hogy az induktív gondolkodás fejlettségi szintje alapvető szerepet tölt be a problémamegoldó folyamatokban, ugyanakkor mind a statikus, mind a dinamikus környezetben mért problémamegoldó képesség jól elkülöníthető az intelligenciától és az induktív gondolkodástól.

A nemzetközi kooperációban kidolgozott harmadik generációs tesztekkel történt vizsgálatok eredményei és az elvégzett elemzések több szempont szerint is hiánypótlóak: 9-19 éves korban vizsgálták a diákok dinamikus problémamegoldó képességének fejlettségi szintjét miközben útelemzések segítségével longitudinális és keresztmetszeti adatok felhasználásával több előrejelző tényező azonosítása történt meg. Miután a számítógép-alapú tesztelés egy jelentős előnye a kontextuális adatok rögzítésének lehetősége, az elemzések során kiemelt figyelmet fordítottunk a logfájelemzésekre. A diákok által alkalmazott stratégiaelemzéseket egy új problémamegoldó stratégiamodellel ismertetése előzte meg, mely lehetővé tette a kontextuális adatok rendszerbe foglalását és további elemzések kivitelezését. A strukturális egyenletek módszeréhez tartozó profilelemzésekkel nemcsak a fejlődés mennyiségi, hanem minőségi változását is detektálni tudtuk és definiálni a különböző tulajdonságokkal jellemezhető diákokat problémamegoldó képességük fejlettségi szintje szerint. Ezek az eredmények új megvilágításba helyezik a korábbi, nemzetközi, főképp egy specifikus stratégia használatának vizsgálatára fókuszáló MicroDYN környezetben történt elemzéseket és azok értelmezését.

Az útelemzések egyrészt alátámasztották a szakirodalom alapján feltételezetteket, miszerint a diákok induktív gondolkodásának fejlettségi szintje jelentős prediktív erővel bír problémamegoldó gondolkodásuk fejlettségi szintjére, másrészt felhívták a figyelmet az iskolai sikeresség és a tudás alkalmazhatósága, a problémamegoldó gondolkodás fejlettségi szintje közötti szakadéokra. Az induktív gondolkodás mellett a gondolkodtató feladatokat és nem a tanultak szó szerinti visszaadását kérő Országos kompetenciamérés matematika tesztjén elért eredmények előrejelző ereje is előtérbe került az elemzések során, amit a diákok neve és angol nyelvtudása követett (utóbbi jelentős mértékben összefügg az induktív gondolkodás fejlettségi szintjével). Az iskolai jegyek és a diszciplináris tudásra építő tesztek ugyanakkor egyáltalán nem jelezték előre a diákok problémamegoldó képességfejlettségi szintjét. Az a kutatási

eredmény, hogy a diákok tanulmányi átlagából, iskolai sikerességéből egyáltalán nem következtethetünk problémamegoldó képességük – beleértve a tudásuk alkalmazhatóságának képességét – fejlettségi szintjére, elgondolkodtató és komoly problémákat vet fel iskolarendszerünk hatékonyságát és működését illetően. A hét év távlatában elvégzett longitudinális elemzések a korábbiakon felül kiemelték a DIFER teszt jó előrejelző hatását is a gondolkodási képesség későbbi fejlettsége tekintetében. Az online tesztekkel történt adatfelvétel további jelentősége, hogy az instrukció meghallgathatóságának lehetőségével a mérés során sikerült teljes mértékben kiküszöbölni a diákok olvasási képessége fejlettségi szintjének teljesítménybefolyásoló hatását, ami a korábbi papíralapú kutatásokban jelentős tényezőnek bizonyult.

A logfájl-elemzések alapján megállapítható, hogy a diákok által alkalmazott stratégia hatékonysága nem minden esetben egyezett meg teljesítményszintjükkel, ugyanakkor a teszt megbízhatósági mutatói jelentős mértékben növekedtek, ha nemcsak a diákok konkrét válaszaira alapoztuk az elemzéseket, hanem figyelembe vettük az általuk alkalmazott stratégia helyességét is. Annak ellenére, hogy a minimálisan komplex rendszerek kiismerése, feltérképezése számos problémamegoldó stratégiával megvalósítható, mégis a szakirodalomban közel egyedülként tárgyalt VOTAT-stratégiákhoz sorolható teljes mértékben elszigetelt változókezelésen alapuló stratégiák alkalmazása bizonyult a leghatékonyabbnak. A többi, elméletileg jó stratégia alkalmazása esetén minden esetben magasabb volt a helytelen értelmezés, a helytelen válaszok aránya, mint a jó megoldásoké. A tudatos stratégiahasználók nagyobb arányban értelmezték helyesen és képezték le jól a rendszerből kinyert információkat, mégsem jelentett egyértelműen helyes megoldást a tudatos stratégiaalkalmazás. Minél bonyolultabb volt a probléma mögött húzódó rendszer, annál inkább elkülönült egymástól a tudatos és kevésbé tudatos problémamegoldók teljesítménye.

A látens profilelemzések alapján az alkalmazott problémafeltérképező stratégia függvényében hat különböző tulajdonságokkal jellemezhető csoportot különíthettünk el egymástól: a legegyszerűbb rendszereket sem kezelő problémamegoldókat, ide sorolhatóak a 6-8. évfolyamos diákok harmada. A legegyszerűbb rendszereket alacsony fokon átlátó és kisebb fokú tanulást mutató problémamegoldókat (16%), az egyszerű problémákat kisebb sikerekkel átlátókat, de a bonyolultabb rendszerekkel nem boldoguló problémamegoldókat (21%). Azon diákok, akik e három csoport valamelyikébe sorolhatóak, alapvetően alacsony szintű explorációs képességekkel rendelkeztek (a diákok 70%-a) és tanulási képességük is kis hatásfokú volt. A diákok 8-10%-a magas teljesítményt mutatott az egyszerű felépítésű problémák megoldásában, ugyanakkor a bonyolultabb rendszereknél már egyértelműen az alulteljesítő problémamegoldók csoportjához tartoztak. Tőlük egyértelműen elkülöníthetőek, bár átlagos teljesítményük tekintetében hasonlóak, a nagyon hatékony tanulási képességekkel rendelkező diákok (6-8%), akik gyorsan megtanulták a rendszer használatát, aminek következtében a teszt végére már a legbonyolultabb problémákat is a szakértő problémamegoldók szintjén oldották meg. A 6-8. évfolyamosok mindössze 17%-át sorolhattuk a szakértő problémamegoldók közé, akik mind az egyszerű, mind a bonyolult problémákat hatékonyan térképezték fel. A tág életkori intervallumot átfogó, ezért több iskolaváltás hatását is vizsgáló elemzések megerősítették azt a feltételezésünket, hogy bár e fontos képesség explicit fejlesztése nem valósul meg a közoktatás folyamán, mégis a szelekció során nagyon lényeges szerepet játszik, sőt a szelekció mértéke erősen magyarázható a diákok problémamegoldó és explorációs képességének fejlettségi szintjével. Például az egyetemista korosztályban, az

érettségit követő szelekció hatására a szakértő problémamegoldók csoportjába sorolható diákok aránya meghaladja a 60%-ot, miközben arányuk a teljes mintában – a fejlődési görbék alapján extrapolálva az eredményeket – 20% körüli. A látens profilelemzések alátámasztották azt az elméletet, miszerint a képességek fejlődése mind mennyiségi, mind minőségi változáson alapul (Csapó, 2003). A technológiaalapú tesztelés és az új elemzési eljárások segítségével mindkettő számszerűsíthető és objektív eszközökkel jellemezhető.

Összességében megállapítható, hogy a technológiaalapú tesztelés és az új módszertani repertoár integrálása olyan új lehetőségeket teremtett a mérés-értékelés terén, amire hagyományos, papíralapú és szemtől szembeni technikák, illetve a klasszikus tesztelmélet alkalmazásával nem volt mód. Ezen eszközök alkalmazásával a korábbinál pontosabbá, objektívabbá, a diákok számára motiválóbbá tehető a tesztelés folyamata. A logfájelemzések segítségével pontosabban rekonstruálható, mit tett, hogyan gondolkodott a diák a tesztelés során, ami új lehetőségeket teremtett a vizsgált jelenségek alaposabb megértése kapcsán. Az értékelés azonnali visszacsatolásának lehetőségét kihasználva a korábbi szummatív dominanciájú megközelítés mellett jelentős hangsúlyt kaphat a diagnosztika, az egyénre szabott, hatékony, tanulást segítő tesztelés. Ezzel megvalósulna a mérés-értékelés átdefiniálása és tanulást segítő funkciójának kihasználása. Alkalmazása a 21. században – egy fejlett oktatási rendszerrel rendelkező ország esetén – elkerülhetetlen.

További kutatási irány a középiskolás diákok problémamegoldó képességszintje fejlődésének pontosabb mennyiségi és minőségi jellemzése. Szintén a további kutatási feladatok közé tartozik a tanulási módszerek és stílusok, valamint a problémák mögött húzódó rendszer megismerése során használt explorációs stratégiák kapcsolatának meghatározása, mely kutatások már folyamatban vannak. Szintén folyamatban vannak a dinamikus problémamegoldó képesség komponensképességeinek teljeskörűbb feltérképezésére irányuló kutatások, melyeket nemzetközi kontextusban is elhelyeztünk. A jelenleg futó nemzetközi összehasonlító kutatássorozatban hasonló mintán összevetjük a kínai és a magyar 5-6. évfolyamos diákok dinamikus problémamegoldó képességének minőségi és mennyiségi jellemzői mellett a dinamikus problémamegoldó képesség szakirodalom alapján feltételezett komponensképességeinek fejlettségi szintjét. Miután a kínai diákok a nemzetközi méréseken (pl.: OECD PISA) az élvonalban helyezkednek el, nemzetközi szinten is értékes kutatási eredményekre számíthatunk az élvonalbeli eredmények okainak, legalábbis egy részének feltárása kapcsán e kutatás keretein belül. Az eredmények alkalmazhatóak lesznek a magyar diákok problémamegoldó képességének fejlesztése kapcsán. Egy további kutatási irány, mely kutatásokat német kollégákkal történő kooperációban valósítjuk meg, a kollaboratív problémamegoldó képesség mérésére alkalmas valid, jó tesztelmélet mutatókkal bíró, azonnali visszacsatolást biztosító teszt kidolgozása, amely lehetővé teszi a kollaboratív problémamegoldó képesség jellemzőinek, illetve fejlődésének feltérképezését. E téren már elkészültek az eDia online mérés-értékelési rendszer azon funkciói, ami lehetővé teszi az ember-ember (*human-to-human*) kooperációt, sőt az első pilot vizsgálatok is lezajlottak már. A kutatás nemzetközi szinten is egyedülálló, ugyanis a kooperatív problémamegoldó képesség vizsgálatát általában ember-számítógép (*human-to-agent*) együttműködésben valósítják meg. További kutatások, valamint új módszertan, elemzései repertoár kidolgozása szükséges a logfájelemzésekben lévő lehetőségek magasabb szintű kihasználásához.



## 7. IRODALOMJEGYZÉK

- ACARA (2012): ACARA National assessment program – ICT literacy Years 6 & 10 Report. Australian Curriculum, Assessment and Reporting Authority, Sydney.
- Adema, J. J. és Gademann, A. J. R. M. (1992): Computerized test construction. In: Wilson, M. (szerk.): *Objective measurement. Theory into practice*. Ablex Publishing Corporation, Norwood, New Jersey. 261–273.
- Adey, P., Csapó, B., Demteriou, A., Hautamäki, J. és Shayer, M. (2007): Can we be intelligent about intelligence? Why education needs the concept of plastic general ability. *Educational Research Review*, **2**. 2. sz. 75–97.
- American Educational Research Association American Psychological Association & National Council on Measurement in Education (1999): *Standards for educational and psychological testing (3rd edition)*. AERA, Washington, DC.
- Applegate, B. (1993): Construction of geometric analogy problems by young children in a computer-based test. *Journal of Educational Computing Research*, **9**. 1. sz. 61–77.
- Artelt, C., Baumert, J., Julius-McElvany, N. és Peschar, J. (2003): *Lerners for life, student approaches to learning. Results from PISA 2000*. OECD, Paris.
- Babad, E., Peer, A. és Hobbs, R. (2012): Media literacy and media bias: Are media literacy students less susceptible to non-verbal judgment biases? *Psychology of Popular Media Culture*, **1**. 97–107.
- Baker, E. L. és Mayer, R. E. (1999): Computer-based assessment of problem solving. *Computer sin Human Behavior*, **15**. 269–282.
- Balázs Ildikó és Zempléni András (2004): A hozottérték-index és a hozzáadott pedagógiai érték számítása a 2003-as kompetenciamérésben. *Új Pedagógiai Szemle*, **12**. 36–50.
- Balázs Ildikó, Felvégi Emese, Rábainé Szabó Annamária és Szepesi Ildikó (2006): *Országos kompetenciamérés 2006. Tartalmi keret*. suliNova Kht., Budapest.
- Balázs Ildikó, Lak Ágnes Rozina, Szabó Vilmos, Vadász Csaba (2013): *Országos kompetenciamérés 2012, Országos jelentés*. Oktatási Hivatal, Budapest.
- Balázs Ildikó, Ostorics László és Szalay Balázs (2007): *PISA 2006. Összefoglaló jelentés. A ma oktatása és a jövő társadalma*. Oktatási Hivatal, Budapest.
- Balázs Ildikó, Ostorics László, Szalay Balázs és Szepesi Ildikó (2010): *PISA 2009 Összefoglaló jelentés. Szövegértés tíz év távlatában*. Oktatási Hivatal, Budapest.
- Balázs Ildikó, Ostorics László, Szalay Balázs, Szepesi Ildikó és Vadász Csaba (2013): *PISA 2012 Összefoglaló jelentés*. Oktatási Hivatal, Budapest.
- Balázs Ildikó, Rábainé Szabó Annamária, Szabó Vilmos és Szepesi Ildikó (2005): A 2004-es Országos kompetenciamérés eredményei. *Új Pedagógiai Szemle*, **55**. 12. sz. 3–21.
- Balázs, I. és Ostorics, L. (2011): *PISA 2009 Digitális szövegértés - olvasás a világhálón*. Oktatási Hivatal, Budapest.
- Báthory Zoltán (1973): *7 standardizált tantárgyteszt*. Országos Pedagógiai Intézet, Budapest.
- Báthory Zoltán (1983): Az iskolai nevelés néhány összetevőjének vizsgálata egy felmérés tükrében. (TOF-80). *Pedagógiai Szemle*, **2**. sz. 135–139.

- Bawden, D. (2008): Origins and concepts of digital literacy. In: Lankshear, C. és Knobel, M. (szerk.): *Digital literacies: concepts, policies and practices*. Peter Lang, New York, NY. 17–32.
- Bebell, D. és Kay, R. (2010): One to one computing: A summary of the quantitative results from the Berkshire wireless learning initiative. *The Journal of Technology, Learning and Assessment*, **9**. 2. sz. <http://napoleon.bc.edu/ojs/index.php/jtla/article/viewFile/1607/1462>
- Bebell, D. és O'Dwyer, L. M. (2010): Educational outcomes and research from 1: 1 computing settings. *Journal of Technology, Learning, and Assessment*, **9**. 1. sz. 5-15.
- Becker, J. (2004): Computergestütztes Adaptives Testen (CAT) von Angst entwickelt auf der Grundlage der Item Response Theorie (IRT). Digitális disszertáció. Freie Universität, Berlin.
- Becker, J. D., Hodge, C. A. és Sepelyak, M. W. (2010): Assessing technology literacy: The case for an authentic, project-based learning approach. [http://genyes.org/media/freeresources/assessing\\_tech\\_literacy\\_whitepaper.pdf](http://genyes.org/media/freeresources/assessing_tech_literacy_whitepaper.pdf)
- Beckmann, J. F. és Guthke, J. (1995): Complex problem solving, intelligence and learning ability. In: Frensch, P. A. és Funke, J. (szerk.): *Complex problem solving. The European perspective*. NJ: Erlbaum, Hillsdale. 3–25.
- Bejar, I. I., Lawless, R., Morley, M. E., Wagner, M. E., Bennett, R. E. és Revuelta, J. (2003): A feasibility study of on-the-fly item generation in adaptive testing. *Journal of Technology, Learning, and Assessment*, **2**. 3. sz. <http://ejournals.bc.edu/ojs/index.php/jtla/article/view/1663>
- Beller, M. (2013): Technologies in large-scale assessments: New directions, challenges, and opportunities. In: von Davier, M., Gonzalez, E., Kirsch, I. és Yamamoto, K. (szerk.): *The role of international large-scale assessments: Perspectives from technology, economy, and educational research*. Springer, Netherlands. 25–45.
- Bennett, R. E. (2002): *Using electronic assessment to measure student performance. The State education standard*. National State Boards of Education, Washington, DC.
- Bennett, R. E. (2003): *Online assessment and the comparability of score meaning*. Educational Testing Service, Princeton, NJ.
- Bennett, R. E., Braswell, J., Oranje, A., Sandene, B., Kaplan, B. és Yan, F. (2008): Does it matter if I take my mathematics test on computer? A second empirical study of mode effects in NAEP. *Journal of Technology, Learning and Assessment*, **6**. 9. sz. 4–38.
- Bennett, R. E., Goodman, M., Hessinger, J., Kahn, H., Ligget, J., Marshall, F. és Zack, J. (1999): Using multimedia in large-scale computer-based testing programs. *Computers in Human Behavior*, **15**. 3–4. sz. 283–294.
- Bennett, R. E., Jenkins, F., Persky, H. és Weiss, A. (2003): Assessing complex problem-solving performances. *Assessment in Education*, 10. sz. 347–359.
- Bennett, R. E., Persky, H., Weiss, A. R. és Jenkins, F. (2007): *Problem solving in technology-rich environments: A report from the NAEP technology-based assessment project (NCES 2007–466)*. National Center for Education Statistics, U.S. Department of Education. Washington, DC.

- Bennett, R. E. (2010): Innovative assessment systems: The role of new technology. Előadás. National Academy of Science Best Practices for State Assessment #2. 2010. április 7, Amerikai Egyesült Államok.
- Benson, N., Hulac, D. M. és Kranzler, J. H. (2010): Independent examination of the Wechsler Adult Intelligence Scale – Fourth Edition: What does the WAIS-IV measure? *Psychological Assessment*, **22**. 1. sz. 121–130.
- Bentler, P. M. (1990): Comparative fit indexes in structural models. *Psychological Bulletin*, **107**. 2. sz. 238–246.
- Binkley, M., Erstad, E., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M. és Rumble, M. (2012): Defining 21st century skills. In: Griffin, P., McGaw, B. és Care, E. (szerk.): *Assessment and Teaching of 21st Century Skills*. Springer, Dordrecht. 17–66.
- Bisanz, J., Bisanz, G. és Korpan, C. A. (1994): Inductive reasoning. In: Sternberg, R. (szerk.): *Thinking and problem solving*. Academic Press, San Diego.
- Bjerkestrand, O. (2009): The European coherent framework of indicators and benchmarks and implications for computer-based assessment. In: Scheuermann, F. és Björnsson, J. (szerk.): *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing*. Office for Official Publications of the European Communities, Luxemburg. 24–29.
- Blazek, N. L. és Forbey, J. D. (2011): A comparison of validity rates between paper and pencil and computerized testing with the MMPI-2. *Assessment*, **18**. 63–66.
- Blazer, C. (2010): Computer-Based Assessments. <http://files.eric.ed.gov/fulltext/ED544707.pdf>.
- Blunch, N. J. (2010): *Introduction to structural equation modelling using SPSS and AMOS*. SAGE, London.
- Bodmann, S. M. és Robinson, D. H. (2004): Speed and performance differences among computer-based and paper-pencil tests. *Journal of Educational Computing Research*, **31**. 1. sz. 51–60.
- Bollen, K. A. (1989): *Structural equations with latent variables*. Wiley, New York.
- Bollen, K. A. és Curran, P. J. (2006): *Latent curve models. A structural equation perspective*. NJ: Wiley, Hoboken.
- Bond, T. és Fox, C. M. (2001): *Applying the Rasch model. Fundamental measurement in the human sciences*. Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey.
- Bond, T. és Fox, C. M. (2015): *Applying the Rasch model. Fundamental measurement in the human sciences*. 3rd edition. Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey.
- Breiter, A., Groß, L. M. és Stauke, E. (2013): Computer-based large-scale assessments in Germany. In: Passey, D., Breiter, A. és Visscher, A. (szerk.): *Next generation of information technology in educational management*. Springer Berlin, Heidelberg. 41–54.
- Bridgeman, B. (2010): Experiences from large-scale computer-based testing in the USA. In: Scheuermann, F. és Björnsson, J. (szerk.): *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing*. European Communities, Brussels. 39–44.

- Bridgeman, B., Lennon, M. L. és Jackenthal, A. (2003): Effects of screen size, screen resolution, and display rate on computer-based test performance. *Applied Measurement in Education*, **16**. 3. sz. 191–205.
- Bridgeman, B., Powers, D., Stone, E. és Mollaun, P. (2012): TOEFL iBT Speaking Test scores as indicators of oral communicative language proficiency. *Language Testing*, **29**. 1. sz. 91–108.
- Brown, T. (2006): CFA with equality constraints, multiple groups, and mean structures. In: Brown, T. (szerk.): *Confirmatory factor analysis for applied research*. Guilford Press, New York. 236–319.
- Browne, M. W. és Cudeck, R. (1993): Alternative ways of assessing model fit. In: Bollen, K. A. és Long, J. S. (szerk.): *Testing structural equation models*. Sage Publications, Thousand Oaks. 136–136.
- Buchner, A. (1995): Basic topics and approaches to the study of complex problem solving. In: Frensch, P. A. és Funke, J. (szerk.): *Complex problem solving: The European perspective*. NJ: Erlbaum, Hillsdale. 27–63.
- Bundy, A. (2004): One essential direction: information literacy, information technology fluency. *Journal of eLiteracy*, **1**. 1. sz. [http://www.jelit.org/6/01/JeLit\\_Paper\\_1.pdf](http://www.jelit.org/6/01/JeLit_Paper_1.pdf).
- Burstein, J., Chodorow, M. és Leacock, C. (2004): Automated essay evaluation: The criterion online service. *AI Magazine*, **25**. 3. sz. 27–36.
- Butcher, J. N. (1987): *Computerized psychological assessment: A practitioner's guide*. Basic Books, New York.
- Byrne, B. M. (2008): Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema*, **20**. 4. sz. 872–882.
- Byrne, B. M. (2010): *Structural equation modeling with AMOS*. Routledge, New York.
- Byrne, B. M. (2012): *Structural equation modeling with MPlus. Basic concepts, applications, and programming*. Routledge, New York, London.
- Byrne, B. M. és Stewart, S. M. (2006): The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling*, **13**. 2. sz. 287–321.
- Carson, K., Gillon, G. és Boustead, T. (2011): Computer-administrated versus paper-based assessment of school-entry phonological awareness ability. *Asia Pacific Journal of Speech, Language and Hearing*, **4**. 2. sz. 85–101.
- Cattel, R. B. és Weiss, R. H. (1980): *Culture fair intelligence test, Scale 3 (CFT3)*. Hogrefe, Göttingen.
- Chi, M. T. H., Glaser, R. és Rees, E. (1982): ‘Expertise in problem solving.’ In: Sternberg, R. S. (szerk.): *Advances in the psychology of human intelligence*. NJ: Erlbaum, Hillsdale. 1–75.
- Choi, S. W. és Tinkler, T. (2002): Evaluating comparability of paper and computer based assessment in a K-12 setting. Előadás. Annual Meeting of the National Council on Measurement in Education. 2002. április 15, New Orleans, Amerikai Egyesült Államok.
- Christakoudis, C., Androulakis, G. S. és Zagouras, C. (2011): Prepare items for large scale computer based assessment: Case study for teachers’ certification on basic computer skills. *Procedia-Social and Behavioral Sciences*, **29**. 1189–1198.

- Cisco, Intel és Microsoft (2009): Transforming education: Assessing and teaching 21st Century skills. <http://atc21s.org/wp-content/uploads/2011/04/Cisco-Intel-Microsoft-Assessment-Call-to-Action.pdf>.
- Clariana, R. és Wallace, P. (2002): Paper-based versus computer-based assessment: key factors associated with the test mode effect. *British Journal of Educational Technology*, **33**. 5. sz. 593–602.
- Cohen, J. (1988): *Statistical power analysis for the behavior sciences*. Erlbaum, Hillsdale.
- Collins, L. M. és Lanza, S. T. (2010): *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. Wiley: New York.
- Computer Science Teachers Association Task Force (2011): *K–12 Computer science standards*. ACM, New York.
- Couse, L. J. és Chen, D. W. (2010): A tablet computer for young children? Exploring its viability for early childhood education. *Journal of Research on Technology in Education*, **43**. 1. sz. 75–96.
- Crook, C. (1992): Young children's skill in using a mouse to control a graphical computer interface. *Computers & Education*, **19**. 3. sz. 199–207.
- Csapó Benő (1993): Tudásszintmérő tesztek. In: Falus Iván (szerk.): *A pedagógiai kutatás módszerei*. Keraban Kiadó, Budapest. 277–316.
- Csapó Benő (1994): Az induktív gondolkodás fejlődése. *Magyar Pedagógia*, 1–2. sz. 53–80.
- Csapó Benő (1998): Az új tudás képződésének eszköze: az induktív gondolkodás. In: Csapó Benő (szerk.): *Az iskolai tudás*. Osiris Kiadó, Budapest. 251–280.
- Csapó Benő (2000): Tudásszintmérő tesztek. In: Falus Iván (szerk.): *A pedagógiai kutatás módszerei*. Műszaki Könyvkiadó, Budapest. 277–316.
- Csapó Benő (2001): Az induktív gondolkodás fejlődésének elemzése országos reprezentatív felmérés alapján. *Magyar Pedagógia*, **101**. 3. sz. 373–391.
- Csapó Benő (2003): *A képességek fejlődése és iskolai fejlesztése*. Akadémiai Kiadó, Budapest.
- Csapó Benő (2005): A komplex problémamegoldás a PISA 2003 vizsgálatban. *Új Pedagógiai Szemle*, 3. sz. 43–52.
- Csapó Benő (2007): Hosszmetszeti felmérések iskolai kontextusban - az első átfogó magyar iskolai longitudinális kutatási program elméleti és módszertani keretei. *Magyar Pedagógia*, **107**. 4. sz. 321–355.
- Csapó Benő (2008): A tanulás és tanítás tudományos megalapozása. In: Fazekas Károly, Köllő János és Varga Júlia (szerk.): *Zöld könyv a magyar közoktatás megújításáért*. Ecostat, Budapest. 217–234.
- Csapó Benő (2014, szerk.): Az érettségi vizsgarendszer megújításának lehetőségei a technológiaalapú tesztelés segítségével. Háttér tanulmány.
- Csapó Benő és Molnár Gyöngyvér (2012): Gondolkodási készségek és képességek. In: Csapó Benő (szerk.): *Mérlegen a magyar iskola*. Nemzeti Tankönyvkiadó, Budapest. 407–440.
- Csapó Benő, Lőrincz András és Molnár Gyöngyvér (2012): Innovative assessment technologies in educational games designed for young students. In: Ifenthaler, D., Eseryel, D. és Ge, X.: *Assessment in game-based learning: Foundations, innovations, and perspectives*. Springer, New York. 235–254.

- Csapó Benő, Molnár Gyöngyvér és Kinyó László (2009): A magyar oktatási rendszer szelektivitása a nemzetközi összehasonlító vizsgálatok eredményeinek tükrében. *Iskolakultúra*, **19**. 3–4. sz. 3–13.
- Csapó Benő, Molnár Gyöngyvér és Nagy József (2015): A DIFER tesztek online változatával végzett mérések tapasztalatai. In: Csapó Benő és Zsolnai Anikó (szerk.): *Online diagnosztikus mérések az iskola kezdő szakaszában*. Oktatáskutató és Fejlesztő Intézet, Budapest. 163–182.
- Csapó Benő, Molnár Gyöngyvér és R. Tóth Krisztina (2008): A papíralapú tesztektől a számítógépes adaptív tesztelésig: a pedagógiai mérés-értékelés technikájának fejlődési tendenciái. *Iskolakultúra*, **18**. 3–4. sz. 3–16.
- Csapó Benő, Molnár Gyöngyvér és R. Tóth Krisztina (2009): Comparing paper-and-pencil and online assessment of reasoning skills: A pilot study for introducing TAO in large-scale assessment in Hungary. In: Scheuermann, F. és Björnsson, J. (szerk.): *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing*. Office for Official Publications of the European Communities, Luxemburg. 113–118.
- Csapó Benő, Molnár Gyöngyvér, Pap-Szigeti Róbert és R. Tóth Krisztina (2009): A mérés-értékelés új tendenciái: a papír- és számítógép-alapú tesztelés összehasonlító vizsgálatai általános iskolás, illetve főiskolás diákok körében. In: Kozma Tamás és Perjés István (szerk.): *Új kutatások a neveléstudományokban*. MTA Pedagógiai Bizottság, Budapest. 99–108.
- Csapó, B., Ainley, J., Bennett, R. E., Latour, T. és Law, N. (2012): Technological issues for computer-based assessment. In: Griffin, P., McGaw, B. és Care, E. (szerk.): *Assessment and teaching of 21st Century skills*. Springer, New York. 143–230.
- Csapó, B., Molnár, Gy. és Nagy, J. (2014): Computer-based assessment of school-readiness and reasoning skills. *Journal of educational psychology*, **106**. 2. sz. 639–650.
- Csapó, B. és Nikolov, M. (2009): The cognitive contribution to the development of proficiency in a foreign language. *Learning and Individual Differences*, **19**. 2. sz. 209–218.
- Csíkos Csaba és B. Németh Mária (2002): A tesztekkel mérhető tudás. In: Csapó Benő (szerk.): *Iskolai tudás*. Második kiadás. Osiris Kiadó, Budapest. 91–122.
- Csíkos Csaba, Molnár Gyöngyvér és Csapó Benő (2015): A matematika online diagnosztikus mérések tartalmi kereteinek elméleti alapjai. In: Csapó Benő, Csíkos Csaba és Molnár Gyöngyvér (szerk.): *A matematikai tudás online diagnosztikus értékelésének tartalmi keretei*. Oktatáskutató és Fejlesztő Intézet, Budapest. 13–26.
- Csíkszentmihályi Mihály (1997): *Flow. Az áramlat: a tökéletes élmény pszichológiája*. Akadémiai Kiadó, Budapest.
- de Konig, E. (2000): *Inductive reasoning in primary education. Measurement, teaching, transfer*. Zeist, Kerckebosch.
- Dikli, S. (2006): An overview of automated scoring of essays. *The Journal of Technology, Learning, and Assessment*, **5**. 1. sz. <http://ejournals.bc.edu/ojs/index.php/jtla/article/view/1640/1489>.
- Dolan, R. P., Burling, K., Harms, M., Strain-Seymour, E., Way, W. és Rose, D. H. (2013): A universal design for learning-based framework for designing accessible technology-enhanced assessments. Research Report.

- [http://images.pearsonclinical.com/images/tmrs/DolanUDL-TEAFramework\\_final3.pdf](http://images.pearsonclinical.com/images/tmrs/DolanUDL-TEAFramework_final3.pdf).
- Donker, A. és Reitsma, P. (2007a): Aiming and clicking in young children's use of the computer mouse. *Computers in Human Behavior*, **23**. 6. sz. 2863–2874.
- Donker, A. és Reitsma, P. (2007b): Drag-and-drop errors in young children's use of the mouse. *Interacting with computers*, **19**. 2. sz. 257–266.
- Donker, A. és Reitsma, P. (2007c): Young children's ability to use a computer mouse. *Computers & Education*, **48**. 4. sz. 602–617.
- Dörner Helga (2007): Kollaboratív tudásépítés számítógéppel segített tanulási környezetben – a tudásépítő interakciók elemzése. MultiMédia az Oktatásban 2007 konferencia, Budapesti Műszaki Főiskola, 2007. augusztus 23-24.
- Dörner, D., Kreuzig, H. W., Reither, F. és Stäudel, T. (1983): *Lohhausen. Vom Umgang mit Unbestimmtheit und Komplexität*. Hans Huber, Bern.
- Duncker, K. (1945): On problem solving. *Psychological Monographs*, **58**. 5. sz. 1–113.
- EETT (online): Enhancing education through technology. <http://www.cde.ca.gov/ls/et/ft/eett.asp>.
- Egan, D. W. és Greeno, J. G. (1974): Theories of rule induction: knowledge acquired in concept learning, serial pattern learning and problem solving. In: Gregg, L.W. (szerk.): *Knowledge and cognition*. John Wiley, New York. 43–104.
- Európa Tanács (2004): Oktatás és képzés 2010 munkaprogram végrehajtása. B munkacsoport: kulcskompetenciák. Implementation of „Education and training 2010” Work Programme. Working Group B „Key Competences”. Key competences for lifelong learning. A European reference framework, November 2004. <http://www.ofi.hu/tudastar/nemzetkozi-kitekintes/egesz-eleten-at-tarto>.
- European Parliament (2006): Recommendation on key competences for lifelong learning. [http://europa.eu/legislation\\_summaries/education\\_training\\_youth/lifelong\\_learning/c11090\\_en.htm](http://europa.eu/legislation_summaries/education_training_youth/lifelong_learning/c11090_en.htm)
- European Commission (2013): Survey of Schools: ICT in education. Benchmarking access, use and attitudes to technology in Europe's schools. Final report. <https://ec.europa.eu/digital-agenda/sites/digital-agenda/files/KK-31-13-401-EN-N.pdf>.
- Falus Iván (1993, szerk.): *A pedagógiai kutatás módszerei*. Keraban Kiadó, Budapest.
- Fan, X. és Sivo, S. A. (2005). Sensitivity of fit indexes to misspecified structural or measurement model components: Rationale of two-index strategy revisited. *Structural Equation Modeling*, **12**. 3. sz. 343–367.
- Farcot, M. és Latour, T. (2008): An open source and large - scale computer based assessment platform: A real winner. In: Scheuermann, F. és Pereira, A. G. (szerk.): *Towards a research agenda on computer - based assessment: Challenges and needs for European Educational Measurement*. European Commission Joint Research Centre, Ispra. 64–67.
- Farcot, M. és Latour, T. (2009): Transitioning to computer-based assessments: A question of costs. In: Scheuermann, F. és Björnsson, J. (szerk.): *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing*. European Communities, Brussels. 108–116.

- Fáyné Dombi Alice, Hódi Ágnes és Kiss Renáta (2016): IKT az óvodában: kihívások és lehetőségek. *Magyar Pedagógia*, **116**. 1. sz. 91–117.
- Fischer, A., Greiff, S. és Funke, J. (2012): The process of solving complex problems. *Journal of Problem Solving*, **4**. 19–42.
- Fraillon, J., Ainley, J., Gebhardt, E. és Schulz, W. (2013): Measuring computer and information literacy across countries. [http://www.iea.nl/fileadmin/user\\_upload/IRC/IRC\\_2013/Papers/IRC-2013\\_Fraillon\\_etal.pdf](http://www.iea.nl/fileadmin/user_upload/IRC/IRC_2013/Papers/IRC-2013_Fraillon_etal.pdf).
- Frensch, P. A. és Funke, J. (1995, szerk.): *Complex problem solving: The European perspective*. Lawrence Erlbaum Associates. Hillsdale, NJ.
- Frey, A. (2007): Adaptives Testen. In: Moosbrugger, H. és Kelava, A. (szerk.): *Testtheorie und Testkonstruktion*. Springer, Berlin, Heidelberg. 261–278.
- Funke, J. (1992): *Wissen über dynamische Systeme: Erwerb, repräsentation und anwendung*. Springer, Heidelberg.
- Funke, J. (2001): Dynamic systems as tools for analysing human judgement. *Thinking and Reasoning*, **7**. 1. sz. 69–89.
- Funke, J. (2010): Complex problem solving: A case for complex cognition? *Cognitive Processing*, **11**. 133–142.
- Funke, J. (2014). Analysis of minimal complex systems and complex problem solving require different forms of causal cognition. *Frontiers in psychology*, **5**. 739.
- Gallagher, A., Bridgeman, B. és Cahalan, C. (2002): The effect of computer-based tests on racial-ethnic and gender groups. *Journal of Educational Measurement*, **39**. 2. sz. 133–147.
- Gamire, E. és Pearson, G. (2006, szerk.): *Tech tally: Approaches to assessing technological literacy*. National Academies Press, Washington, DC.
- Gardner, P. H. és Berry, D. C. (1995): The effect of different forms of advice on the control of a simulated complex system. *Applied Cognitive Psychology*, **9**. 7. sz. 55–79.
- Gerry, W. (2008): ICT trends in education. Teaching and learning and leadership digital learning research. Australian Council for Educational Research. [http://research.acer.edu.au/cgi/viewcontent.cgi?article=1001&context=digital\\_learning](http://research.acer.edu.au/cgi/viewcontent.cgi?article=1001&context=digital_learning).
- Gierl, M. J. és Haladyna, T. M. (2013, szerk.): *Automatic item generation: Theory and practice*. Routledge, New York and London.
- Gilhooly, K. J. (1982): *Thinking: Directed, undirected and creative*. Academic Press, London.
- Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H. és Klieme, E. (2014): The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology*, **106**. 3. sz. 608–626.
- Goldman, S. R., Lawless, K., Pellegrino, J. és Gomez, K. (2012): A technology for assessing multiple source comprehension: An essential skill of the 21st Century. In: Mayrath, M. C., Clarke-Midura, J., Robinson, D. H. és Schraw, G. (szerk.): *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research*. Information Age Publishing, Charlotte, NC. 173–210.
- Gonzalez, C., Thomas, R. P. és Vanyukov, P. (2005): The relationships between cognitive ability and dynamic decision making. *Intelligence*, **33**. 2. sz. 169–186.



- Gordon Győri János (2002): A gondolkodási képességek fejlesztésének pedagógiája Szingapúrban. *Magyar Pedagógia*, **102**. 2. sz. 203–229.
- Graesser, A. C. (2012): Foreword. In: Mayrath, M. C., Clarke-Midura, J., Robinson, D. H. és Schraw, G. (szerk.): *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research*. Information Age Publishing, Charlotte, NC. vii-ix.
- Greiff, S. (2012): From interactive to collaborative problem solving: Current issues in the Programme for International Student Assessment. *Review of Psychology*, **19**. 2. sz. 111–121.
- Greiff, S. (2015): Educating students towards becoming proficient problem solvers: Initiating an important discussion around 21st century education. Előadás. 7th Szeged Workshop on Educational Evaluation. 2015. április 27-28.
- Greiff, S. és Funke, J. (2010): Systematische Erforschung komplexer Problemlösefähigkeit anhand minimal komplexer Systeme. *Zeitschrift für Pädagogik*, **56**. 216–227.
- Greiff, S., Holt, D. V. és Funke, J. (2013): Perspectives on problem solving in cognitive research and educational assessment: Analytical, interactive, and collaborative problem solving. *Journal of Problem Solving*, **5**. 71–91.
- Greiff, S., Wüstenberg, S. és Avvisati, F. (2015). Computer-generated log-file analyses as a window into students' minds? A showcase study based on the PISA 2012 assessment of problem solving. *Computers & Education*, **91**. 92–105.
- Greiff, S., Wüstenberg, S., Csapó, B., Demetriou, A. Hautamäki, H., Graesser, A. C. és Martin, R. (2014): Domain-general problem solving skills and education in the 21st century. *Educational Research Review*, **13**. 74–83.
- Greiff, S., Wüstenberg, S. és Funke, J. (2012): Dynamic problem solving: A new assessment perspective. *Applied Psychological Measurement*, **36**. 3. sz. 189–213.
- Greiff, S., Wüstenberg, S., Molnár, G., Fischer, A., Funke, J. és Csapó, B. (2013). Complex problem solving in educational contexts – Something beyond g: Concept, assessment, measurement invariance, and construct validity. *Journal of Educational Psychology*, **105**. 2. sz. 364–379.
- Griffin, B., McGaw, B. és Care, E. (2012, szerk.): *Assessment and teaching 21st century skills*. Springer, Dordrecht, Germany.
- Griffin, P. (1999): *Item response modelling: An introduction to the Rasch Model*. Assessment Research Centre Faculty of Education, The University of Melbourne.
- Grünzweil, B. és Haller, M. (2009): Analysing interaction techniques using mouse and keyboard for preschool children. In: Holzinger, A. és Miesenberger, K. (szerk.): *HCI and usability for e-inclusion*. Springer, Berlin Heidelberg, New York. 448–456.
- Hahne, J. (2008): Analyzing position effects within reasoning items using the LLTM for structurally incomplete data. *Psychology Science Quarterly*, **50**. 3. sz. 379–390.
- Hamers, J. H. M., De Koning, E. és Sijtsma, K. (2000): Inductive reasoning in the third grade: Intervention promises and constraints. *Contemporary Educational Psychology*, **23**. 132–148.
- Hawkes, B. (2012): The next generation of assessments: Simulations. <http://www.kenexa.com/Portals/0/Downloads/The%20Next%20Generation%20of%20Assessments.pdf>.

- Herzog Csilla (2012): A médiaműveltség és a médiahasználat vizsgálata 14-18 éves tanulók körében. PhD értekezés. Szeged.
- Herzog Csilla és Racskó Réka (2013): A 14-18 éves tanulók médiatudatosságának empirikus vizsgálata és fontosabb eredményei. In: Karlovitz János Tibor (szerk): *Tanulmányok az emberi gondolkodás tárgykörében*. International Research Institute, Komárno. 12–22.
- Hobbs, R. (2011): The state of media literacy: A response to Potter. *Journal of Broadcasting and Electronic Media*, **55**. 419–430.
- Hódi Ágnes és R. Tóth Krisztina (2009): Olvasási képesség mérése számítógépes környezetben. IX. Országos Neveléstudományi Konferencia, Veszprém, 2009. november 19-21. 96.
- Horkay, N., Bennett, R. E., Allen, N., Kaplan, B. és Yan, F. (2006): Does it matter if I take my writing test on computer? An empirical study of mode effects in NAEP. *Journal of Technology, Learning, and Assessment*, **5**. 2. sz. <http://ejournals.bc.edu/ojs/index.php/jtla/issue/view/180>.
- Horváth György (1997): *A modern tesztmodellek alkalmazása*. Akadémiai Kiadó, Budapest.
- Hourcade, J. P., Bederson, B. B., Druin, A. és Guimbretiére, F. (2004): Differences in pointing task performance between preschool children and adults using mice. *ACM Transactions on Computer-Human Interaction (TOCHI)*, **11**. 4. sz. 357–386.
- Hülber László (2012): A papír és a számítógép-alapú tesztelés összehasonlító vizsgálata különböző item paraméterek mentén. *Iskolakultúra*, **22**. 12. sz. 13–25.
- Hülber László és Molnár Gyöngyvér (2013): Papír és számítógép-alapú tesztelés nagymintás összehasonlító vizsgálata matematika területén, 1-6. évfolyamon. *Magyar Pedagógia*, **113**. 4. sz. 243–263.
- Hunya Márta (2011): Az eLEMÉR keretrendszeréről. <http://ikt.ofi.hu/ikt-onertekelo-keretrendszer/keretrendszerrol>.
- Hunya Márta (2013a): eLEMÉRÉS 2013. [http://ikt.ofi.hu/ikt/wp-content/uploads/MI\\_eLEMERES\\_2013.pdf](http://ikt.ofi.hu/ikt/wp-content/uploads/MI_eLEMERES_2013.pdf).
- Hunya Márta (2013b): IKT1-felmérés az európai iskolákban. [http://essie.eun.org/c/document\\_library/get\\_file?uuid=1d49031b-7457-4c46-b49a-c8923a26c3f9&groupId=21279](http://essie.eun.org/c/document_library/get_file?uuid=1d49031b-7457-4c46-b49a-c8923a26c3f9&groupId=21279).
- Hunya Márta (2015): eLEMÉRÉS 2011-2015. Háttér tanulmány. [http://ofi.hu/sites/default/files/attachments/elemeres\\_2015.pdf](http://ofi.hu/sites/default/files/attachments/elemeres_2015.pdf)
- Hunya Márta, Dancsó Tünde és Tartsayné Németh Nóra (2006): Informatikai eszközök használata a tanítási órákon. *Új Pedagógiai Szemle*, 7–8. sz. <http://epa.oszk.hu/00000/00035/00105/2006-07-in-Tobbek-Informatikai.html>.
- Hunya Márta, Körösné dr. Mikis Márta, Tartsayné Németh Dóra és Tibor Éva (2010): eLEMÉR gyorsjelentése az informatikai eszközök iskolafejlesztő célú alkalmazásáról. OFI, Budapest. [http://ikt.ofi.hu/ikt/wp-content/uploads/elemer\\_gyorsjelentes\\_probameresrol.pdf](http://ikt.ofi.hu/ikt/wp-content/uploads/elemer_gyorsjelentes_probameresrol.pdf).
- Hunya Márta, Körösné dr. Mikis Márta, Tartsayné Németh Nóra és Tibor Éva (2011): Gyorsjelentés az informatikai eszközök iskolafejlesztő célú alkalmazásának országos helyzetéről. <http://ikt.ofi.hu/ikt-tudastar/projekt-tanulmanyok/gyorsjelentes-2011>.

- Inkpen, K. M. (2001): Drag-and-drop versus point-and-click mouse interaction styles for children. *ACM Transactions on Computer-Human Interaction (TOCHI)*, **8**. 1. sz. 1–33.
- ITTK (2007): *Magyar Információs Társadalom Jelentés 1998-2008. – Jelentés az elmúlt évtizedről*. ITTK, Budapest.
- ITU (2015): *Measuring the Information Society Report 2015*. International Telecommunication Union, Geneva, Switzerland. <http://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2015/MISR2015-w5.pdf>
- Jeni, L. A., Lőrincz, A., Nagy, T., Palotai, Z., Sebők, J., Szabó, Z. és Takács, D. (2012): 3D shape estimation in video sequences provides high precision evaluation of facial expressions. *Image and Vision Computing*, <http://nipg.inf.elte.hu/szzoli/publications/jeni12shape.pdf>.
- Jodoin, M. Zenisky A. és Hambleton, R. K. (2006): Comparison of the psychometric properties of several computer-based test designs for credentialing exams with multiple purposes. *Applied Measurement in Education*, **19**. 3. sz. 203–220.
- Johnson, M. és Green, S. (2006): On-line mathematics assessment: The impact of mode on performance and question answering strategies. *Journal of Technology, Learning and Assessment*, **4**. 5. sz. 4–33.
- Johnson-Laird, P. N. (1983): *Mental models: Toward a cognitive science of language, inference and consciousness*. Harvard University Press, Cambridge.
- Joiner, R., Messer, D., Light, P. és Littleton, K. (1998): It is best to point for young children: a comparison of children's pointing and dragging. *Computers in Human Behaviour*, **14**. 3. sz. 513–529.
- Jonassen, D. H. (1997): Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *Educational Technology: Research and Development*, **45**. 1. sz. 65–94.
- Jonassen, D. H. és Grabowski, B. L. (1993): *Handbook of individual differences, learning, and instruction*. Lawrence Erlbaum, Mahwah, NJ.
- Józsa Krisztián és Zentai Gabriella (2007): Hátrányos helyzetű óvodások játékos fejlesztése a DIFER Programcsomag alapján. *Új Pedagógiai Szemle*, 5. sz. 3–17.
- Jude, N. (2006): IRT-Scalierung mit ConQuest. Workshop für das Nachwuchsnetzwerk Deutschdidaktik. DIPF. [http://www.symposion-deutschdidaktik.de/fileadmin/template/download/aktivitaeten/nachwuchsnetzwerk/Folien\\_Jude.pdf](http://www.symposion-deutschdidaktik.de/fileadmin/template/download/aktivitaeten/nachwuchsnetzwerk/Folien_Jude.pdf).
- Jurecka, A. és Hartig, J. (2007). Computer- und Netzbasiertes Assessment. In: Hartig, J. és Klieme, E. (szerk.): *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik*. Bundesministerium für Bildung und Forschung, Bonn, Berlin. 37–48
- Kampylis, P., Law, N., Punie, Y., Bocconi, S., Brečko, B., Han, S., Looi, C-K. és Miyake, N. (2013): *ICT-enabled innovation for learning in Europe and Asia. Exploring conditions for sustainability, scalability and impact at system level*. European Commission Joint Research Centre Institute for Prospective Technological Studies. Seville. <http://ftp.jrc.es/EURdoc/JRC83503.pdf>.
- Kárpáti, A. és Horváth, Á. (2009): National policies and practices on ICT in education in Hungary. In: Plomp, T., Law, N., Anderson, R. és Quale, A. (szerk.): *Cross-*

- national ICT policies and practices in education*. Information Age Publishing, Charlotte, NC, USA, 349–368.
- Katz, I. R. és Macklin, A. S. (2007): Information and communication technology (ICT) literacy: Integration and assessment in higher education. *Systemics, Cybernetics and Informatics*, **5**. 4. sz. 50–55.
- KCC (2014): Annual report 2013. Korea Communications Commission. Korea. <http://eng.kcc.go.kr/download.do?fileSeq=42390>
- Kettler, R. J. (2011): Computer-based screening for the new modified alternate assessment. *Journal of Psychoeducational Assessment*, **1**. 29. sz. 3–13.
- Kies, S. M., Williams, B. D. és Freund, G. G. (2006): Gender plays no role in students ability to perform on computer-based examinations. *BMC Medical Education*, **6**. 57. sz. <http://www.biomedcentral.com/1472-6920/6/57>.
- Kikis, K. (2010): Reflections on paper-and-pencil tests to eAssessments: Narrow and broadband paths to 21st Century challenges. In: Scheuermann, F. és Bjornsson, J. (szerk.): *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing*. European Communities, Brussels. 99–103.
- Kim, D. és Huynh, H. (2006): Comparison of student performance between paper-and-pencil and computer-based testing in four content areas. Előadás, Annual meeting of National Council on Measurement in Education, 2006. április, San Francisco, Amerikai Egyesült Államok.
- Kim, D. és Huynh, H. (2007): Comparability of computer and paper-and-pencil versions of Algebra and Biology assessments. *Journal of Technology, Learning and Assessment*, **6**. 4. sz. 4–29.
- Kingery, D. és Furuta, R. (1997): Skimming electronic newspaper headlines: A study of typeface, point size, screen resolution, and monitor size. *Information Processing and Management*, **33**. 685–696.
- Kingston, N. M. (2009): Comparability of computer- and paper-administered multiple-choice tests for K-12 populations: A synthesis. *Applied Measurement in Education*, **22**. 1. sz. 22–37.
- Kiss Árpád (1960a): Iskolai tanulóink tudásszintjének vizsgálata. 1. rész. *Pedagógiai Szemle*, **3**. sz. 194–206.
- Kiss Árpád (1960b): Iskolai tanulóink tudásszintjének vizsgálata. 2. rész. *Pedagógiai Szemle*, **7–8**. sz. 585–593.
- Kiss Árpád (1960c): Iskolai tanulóink tudásszintjének vizsgálata. 3. rész. *Pedagógiai Szemle*, **9**. sz. 775–784.
- Kiss Árpád (1961): Iskolai tanulóink tudásszintjének vizsgálata. 4. rész. *Pedagógiai Szemle*, **7–8**. sz. 600–613.
- Kiss Árpád (1969): *Műveltség és iskola*. Akadémiai Kiadó, Budapest.
- Kiss Árpád (1978): *Mérés, értékelés, osztályozás*. Tankönyvkiadó, Budapest.
- Klauer, K. J. (1989): *Denktraining für Kinder I*. Göttingen, Hogrefe.
- Klauer, K. J. (1990): Paradigmatic teaching of inductive thinking. In: Mandl, H., De Corte, E., Bennett, S. N. és Friedrich, H. F. (szerk.): *Learning and instruction. European research in an international context. Analysis of complex skills and complex knowledge domains*. Pergamon Press, Oxford. 23–45.

- Klauer, K. J. (1991): *Denktraining für Kinder II*. Hogrefe, Göttingen.
- Klauer, K. J. (1992): «Bottom up» oder «top down»? Über die Transferwirkungen zweier Strategien zum Training des induktiven Denkens. *Sprache & Kognition*, **11**. 2. sz. 91–103.
- Klauer, K. J. (1993): *Denktraining für Jugendliche*. Hogrefe, Göttingen.
- Klauer, K. J. (1996): Begünstigt induktives Denken das Lösen komplexer Probleme? Experimentellen Studien zu Leutners sahel-Problem. *Zeitschrift für Experimentelle Psychologie*, **43**. 1. sz. 361–366.
- Klauer, K. J. (1997): A tanulás és a kognitív képességek fejlesztése. *Iskolakultúra*, **7**. 12. sz. 85–92.
- Klieme, E. (2004): Assessment of cross-curricular problem-solving competencies. In: Moskowitz, J. H. és Stephens, M. (szerk.): *Comparing learning outcomes. Assessments and international education policy*. Routledge Falmer, London. 81–107.
- Koltay Tibor (2009): Médiaműveltség, média-írástudás, digitális írástudás. [http://www.mediakutato.hu/cikk/2009\\_04\\_tel/08\\_mediamuveltség\\_digitalis\\_irastudas](http://www.mediakutato.hu/cikk/2009_04_tel/08_mediamuveltség_digitalis_irastudas).
- Koltay Tibor (2010): Az új média és az írástudás új formái. *Magyar Pedagógia*, **110**. 4. sz. 301–309.
- Kontra József (1996): A probléma és problémamegoldás. *Magyar Pedagógia*, **96**. 4. sz. 341–365.
- Koong, C. S. és Wu, C. Y. (2011): The applicability of interactive item templates in varied knowledge types. *Computers & Education*, **56**. 3. sz. 781–801.  
<http://www.sciencedirect.com/science/article/pii/S0360131510003064>.
- Korom Erzsébet, Molnár Gyöngyvér és Csapó Benő (2015): A természettudományi online diagnosztikus mérések tartalmi kereteinek elméleti háttere. In: Csapó Benő, Korom Erzsébet és Molnár Gyöngyvér (szerk.): *A természettudományi tudás online diagnosztikus értékelésének tartalmi keretei*. Oktatókutató és Fejlesztő Intézet, Budapest. 13–30.
- Kozma, R. B. (2008): 'Comparative analysis of policies for ICT in education'. In: Voogt, J. és Knezek, G. (szerk.): *International handbook on information technology in primary and secondary education*. Springer, New York. 1083–1096.
- Kröner, S., Plass, J. L. és Leutner, D. (2005): Intelligence assessment with computer simulations. *Intelligence*, **33**. 4. sz. 347–368.
- Kuo, C. Y. és Wu, H. K. (2013): Toward an integrated model for designing assessment systems: An analysis of the current status of computer-based assessments in science. *Computers & Education*, 68. sz. 388–403.
- Kyllonen, P. (2009): New constructs, methods and directions for computer-based assessment. In: Sheuermann, F. és Björnsson, J. (szerk.): *The transition to computer-based assessment. New approaches to skills assessment and implications for large-scale testing*. Office for Official Publications of the European Communities, Luxemburg. 151–156.
- Lane, A. E. és Ziviani, J. M. (2010): Factors influencing skilled use of the computer mouse by school-aged children. *Computers & Education*, **55**. 3. sz. 1112–1122.

- Lane, A. és Ziviani, J. (2002): Enabling computer access: Introduction to the test of mouse proficiency. *OTJR: Occupation, Participation and Health*, **22**. 3. sz. 111–118.
- Latour, T. és Farcot, M. (2008): An open source and large-scale computer-based assessment platform: A real winner. In: Scheuermann, F. és Pereira, A. G. (szerk.): *Towards a research agenda on computer-based assessment. Challenges and needs for European educational measurement*. Office for Official Publications of the European Communities, Luxemburg. 64–67.
- Latour, T. és Martin, R. (2007): TAO, An open and versatile computer-based assessment platform based on semantic web technology. *ECRIM*, <https://www.tao.lu/downloads/publications/TAO-ErcimNews71-Oct2007.pdf>.  
Letöltés ideje: 2010. november 10.
- Law, N., Lee, Y. és Yuen, H. K. (2009): The impact of ICT in education policies on teacher practices and student outcomes in Hong Kong. In: Scheuermann, F. és Pedró, F. (szerk.): *Assessing the effects of ICT in Education - indicators, criteria and benchmarks for international comparisons*. Publications Office of the European Union, Luxembourg. 143–164.
- Law, N., Pelgrum, W. J. és Plomp, T. (2008, szerk.): *Pedagogy and ICT use in schools around the world: findings from the IEA SITES 2006 study*. Comparative Education Research Center, Hong Kong.
- Leeson, H. V. (2006): The mode effect: A literature review of human and technological issues in computerized testing. *International Journal of Testing*, **6**. 1. sz. 1–24.
- Lévai Dóra (2013): A digitális állampolgárság és digitális műveltség kompetenciája a pedagógus tevékenységéhez kapcsolódóan. *Oktatás-Informatika*, 1-2. sz. <http://www.oktatas-informatika.hu/2013/11/levai-dora-a-digitalis-allampolgarsag-es-digitalis-muveltseg-kompetenciaja-a-pedagogus-tevekenysegehez-kapcsolodoan/>.
- Lipponen, L. (2002): Exploring foundations for computer-supported collaborative learning. In: Stahl, G. (szerk.): *Proceedings of computer supported collaborative learning*. Boulder, CO. 72–81.
- Lissitz, R. W. és Jiao, H. (2012, szerk.): *Computers and their impact on state assessment: Recent history and predictions for the future*. Information Age, Charlotte. 25–53.
- Lőrincz, A., Molnár, G., Jeni, L., Tözsér, Z., Rausch, A., Cohn, J. F. és Csapó, B. (2013): Towards entertaining and efficient educational games. NIPS 2013, Data Driven Education, Nevada, USA, 2013. december 05-10.
- Lottridge, S. M., Nicewander, W. A., Schulz, E. M. és Mitzel, H. C. (2010): Comparability of paper-based and computer-based tests: A review of the methodology. In: Winter, P. C. (szerk.): *Evaluating the comparability of scores from achievement test variations*. Council of Chief State School Officers, Washington. 119–152.
- Luecht, R. M. és Nungester, R. J. (1998): Some practical examples of computer-adaptive sequential testing. *Journal of Educational Measurement*, **35**. 229–249.
- MacCann, R. (2006): The equivalence of online and traditional testing for different subpopulations and item types. *British Journal of Educational Technology*, **37**. 1. sz. 79–91.
- Macdonald, J. (2003): Assessing online collaborative learning: process and product. *Computers & Education*, **40**. 4. sz. 377–391.

- Magyar Andrea (2013): Többszakaszos adaptív tesztek felépítése, működése. *Oktatás-Informatika*, 1-2. sz. <http://www.oktatas-informatika.hu/2013/11/magyar-andrea-tobbszakaszos-adaptiv-tesztek-felepitese-mukodese/>.
- Magyar Andrea (2014a): Adaptív tesztek készítésének folyamata. *Iskolakultúra*, **24.** 4. sz. 26–33.
- Magyar Andrea (2014b): Számítógép-alapú adaptív és lineáris tesztek összehasonlító hatékonyságvizsgálata. PhD disszertáció, Szegedi Tudományegyetem, Neveléstudományi Doktori Iskola.
- Magyar Andrea és Molnár Gyöngyvér (2015): Az item pozíciós hatás vizsgálata BIBD klaszterdesign esetén. XV. Országos Neveléstudományi Konferencia, Budapest. 2015. november 19-21. 262.
- Magyar Andrea és Molnár Gyöngyvér (2013): Adaptív és rögzített formátumú tesztek alkalmazásának összehasonlító hatékonyságvizsgálata. *Magyar Pedagógia*, 3. sz. 181–193.
- Magyar Andrea és Molnár Gyöngyvér (2014): A szóolvasási készség adaptív mérését lehetővé tevő online tesztrendszer kidolgozása. *Magyar Pedagógia*, 4. sz. 259–279.
- Magyar Andrea és Molnár Gyöngyvér (2015): A szóolvasási készség online adaptív mérésének hatékonyságvizsgálata. *Magyar Pedagógia*, **115.** 4. sz. 403–428.
- Magyar Andrea: (2012): Számítógépes adaptív tesztelés. *Iskolakultúra*, **22.** 6. sz. 52–60.
- Markman, A. B. (1999): *Knowledge representation*. Erlbaum, Mahwah, NJ.
- Martin, R. (2010): Utilising the potential of computer delivered surveys in assessing scientific literacy. In: Scheuermann, F. és Björnsson, J. (szerk.): *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing*. European Communities, Brussels. 172–177.
- Martin, R. és Binkley, M. (2009): Gender differences in cognitive tests: A consequence of gender-dependent preferences for specific information presentation formats? In: Scheuermann, F. és Björnsson, J. (szerk.): *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing*. Office for Official Publications of the European Communities, Luxembourg. 75–82.
- Martins, N. (2010): Measurement model equivalence in web- and paper-based surveys. *Southern African Business Review*, **14.** 3. sz. 77–107.
- Mayer, R. E. (1998): Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science*, **26.** 49–63.
- Mayrath, M. C., Clarke-Midura, J. és Robinson, D. (2012b): *Technology-based Assessments for 21st Century skills: Theoretical and practical implications from modern research*. Information Age, Charlotte.
- Mayrath, M. C., Clarke-Midura, J. és Robinson, D. H. (2012a): Introduction to technology-based assessments for 21st Century skills. In: Mayrath, M. C., Clarke-Midura, J., Robinson, D. H. és Schraw, G. (szerk.): *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research*. Information Age Publishing, Charlotte, NC. 1–12.

- McGrew, K. S. (2009): CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, **37**. 1–10.
- Meijer, R. (2010): Transition to computer-based assessment: Motivations and considerations. In: Scheuermann, F. és Bjornsson, J. (szerk.): *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing*. European Communities, Brussels. 104–107.
- Meredith, W. (1993): Measurement invariance, factor analyses and factorial invariance. *Psychometrika*, **58**. 4. 525–543.
- Milfont, T. L. és Fischer, R. (2010): Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, **3**. 1. sz. 111–121.
- Mislevy, R. J., Behrens, J. T., Dicerbo, K. E. és Levy, R. (2012): Design and discovery in educational assessment: Evidence-centered design, psychometrics, and educational data mining. *Journal of Educational Data Mining*, **4**. 1. sz. 11–48.
- Mitchell, T., Russel, T., Broomhead, P. és Aldridge, N. (2002). Towards robust computerized marking of free-text responses. In: Danson, M. (szerk.): *Proceedings of the Sixth International Computer Assisted Assessment Conference*. Loughboroug University, Loughborouh, UK. [https://dspace.lboro.ac.uk/dspace-jspui/bitstream/2134/1884/1/Mitchell\\_t1.pdf](https://dspace.lboro.ac.uk/dspace-jspui/bitstream/2134/1884/1/Mitchell_t1.pdf)
- Moe, E. (2008): Introducing large scale computer-based testing of English – experiences and future challenges. Presented at international research workshop. The transition to computer-based assessment, lessons learned from the PISA 2006 Computer Based Assessment of Science (CBAS) and implications for large scale testing, Reykjavik, Iceland, September 29.–October 1.
- Moe, E. (2010): Introducing large-scale computerized assessment – Lessons learned and future challenges. In: Scheuermann, F. és Bjornsson, J. (szerk.): *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing*. European Communities, Brussels. 51–56.
- Molnár Gyöngyvér (2001): Az életszerű feladat-helyzetekben történő problémamegoldás vizsgálata. *Magyar Pedagógia*, **101**. 3. sz. 347–372.
- Molnár Gyöngyvér (2002): Komplex problémamegoldás vizsgálata 9–17 évesek körében. *Magyar Pedagógia*, **102**. 2. sz. 231–264.
- Molnár Gyöngyvér (2003): Az ismeretek alkalmazásának vizsgálata modern tesztelméleti (IRT) eszközökkel. *Magyar Pedagógia*, **103**. 4. sz. 423–446.
- Molnár Gyöngyvér (2004): Hátrányos helyzetű diákok problémamegoldó gondolkodásának fejlettsége. *Magyar Pedagógia*, **104**. 3. sz. 319–338.
- Molnár Gyöngyvér (2005): Az objektív mérés megvalósításának lehetősége: a Rasch-modell. *Iskolakultúra*, **15**. 3. sz. 71–80.
- Molnár Gyöngyvér (2006a): A Rasch-modell alkalmazása a társadalomtudományi kutatásokban. *Iskolakultúra*, **16**. 12. sz. 99–113.
- Molnár Gyöngyvér (2006b): Az ismeretek alkalmazhatóságának korlátai: komplex problémamegoldó gondolkodás fejlettsége 7. és 11. évfolyamon. *Magyar Pedagógia*, **106**. 4. sz. 329–344.



- Molnár Gyöngyvér (2006c): Az induktív gondolkodás fejlesztése kisiskolás korban. *Magyar Pedagógia*, **106**. 1. sz. 63–80.
- Molnár Gyöngyvér (2007): Hátrányos helyzetű diákok problémamegoldó gondolkodásának longitudinális követése. *Magyar Pedagógia*, **107**. 4. sz. 277–293.
- Molnár Gyöngyvér (2008a): A Rasch modell kiterjesztése nem dichotóm adatok elemzésére: a rangskálás és parciális kredit modell. *Iskolakultúra*, **18**. 1. sz. 66–77.
- Molnár Gyöngyvér (2008b): Értékelés-számítógép-alapú tesztelés, online tesztkörnyezet, adaptív tesztelés. In: Kárpáti Andrea, Molnár Gyöngyvér, Tóth Péter és Főző Attila (szerk.): *A 21. század iskolája*. Nemzeti Tankönyvkiadó, Budapest. 47–50.
- Molnár Gyöngyvér (2008c): Kisiskolások induktív gondolkodásának játékos fejlesztése. *Új Pedagógiai Szemle*, 5. sz. 51–64.
- Molnár Gyöngyvér (2009a): Papír- és számítógép-alapú tesztelés összehasonlító vizsgálata problémamegoldó környezetben. IX. Országos Neveléstudományi Konferencia, Veszprém, 2009. november 19-21. 95.
- Molnár Gyöngyvér (2009b): Kisiskolás diákok számára kidolgozott induktív gondolkodás fejlesztő program hosszabb távú hatása. In: Perjés István és Kozma Tamás (szerk.): *Új kutatások a neveléstudományokban. Hatékony tudomány, pedagógiai kultúra, sikeres iskola*. Magyar Tudományos Akadémia, Budapest. 2009. 118-129.
- Molnár Gyöngyvér (2010a): Technológiaalapú mérés-értékelés hazai és nemzetközi implementációi. *Iskolakultúra*, **20**. 7–8. sz. 22–34.
- Molnár Gyöngyvér (2010b): Papír- és számítógép-alapú tesztelés összehasonlító vizsgálata problémamegoldó környezetben. In: Perjés István és Kozma Tamás: *Új kutatások a neveléstudományokban*. Aula Kiadó, Corvinus Egyetem, Budapest. 135–144.
- Molnár Gyöngyvér (2011a): Az információs-kommunikációs technológiák hatása a tanulásra és oktatásra. *Magyar Tudomány*, 9. sz. 1038–1047.
- Molnár Gyöngyvér (2011b): Playful fostering of 6- to 8-year-old students' inductive reasoning. *Thinking skills and Creativity*, **6**. 2. sz. 91–99.
- Molnár Gyöngyvér (2012): A problémamegoldó gondolkodás fejlődése: az intelligencia és szocioökonómiai háttér befolyásoló hatása 3-11. évfolyamon. *Magyar Pedagógia*, **112**. 1. sz. 41–58.
- Molnár Gyöngyvér (2013a): Területspecifikus komplex problémamegoldó gondolkodás fejlődése. In: Molnár Gyöngyvér és Korom Erzsébet (szerk.): *Az iskolai sikerességet befolyásoló kognitív és affektív tényezők értékelése*. Nemzedékek Tudása Tankönyvkiadó, Budapest. 161–180.
- Molnár Gyöngyvér (2013b): *A Rasch modell alkalmazási lehetőségei az empirikus kutatások gyakorlatában*. Gondolat Kiadó, Budapest.
- Molnár Gyöngyvér (2013c): Mindennapi helyzetekben alkalmazott problémamegoldó stratégiák változása. *Iskolakultúra*, **23**. 7-8. sz. 31–43.
- Molnár Gyöngyvér (2015a): Az óvoda és iskola feladatai az értelmi képességek fejlesztése terén. In: Kónyáné Tóth Mária és Molnár Csaba (szerk.): *Tartalmi és szervezeti változások a köznevelésben*. Suliszerviz Oktatási és Szakértői Iroda, Suliszerviz Pedagógiai Intézet, Debrecen. 179–190.

- Molnár Gyöngyvér (2015b): A képességmérés dilemmái: a diagnosztikus mérések (eDia) szerepe és helye a magyar közoktatásban. *Génusz Műhely Kiadványok*, 2. sz. 16–29.
- Molnár Gyöngyvér (2015c): Bevezető. In: Molnár Gyöngyvér, Papp Zoltán, Makay Géza és Ancsin Gábor: *eDia 2.3 Online mérési platform – feladatfelviteli kézikönyv*. SZTE Oktatáselméleti Kutatócsoport, Szeged. 5–7.
- Molnár Gyöngyvér és Csapó Benő (2003): A képességek fejlődésének logisztikus modellezése. *Iskolakultúra*, **13**. 2. sz. 57–69.
- Molnár Gyöngyvér és Csapó Benő (2011): Az induktív gondolkodás fejlődése és az átlagos fejlettségi szint időbeli változásai 1-11. évfolyamon országos reprezentatív minták alapján. *Magyar Pedagógia*, **111**. 2. sz. 127–140.
- Molnár Gyöngyvér és Csapó Benő (2013): Az eDia online diagnosztikus mérési rendszer. XI. Pedagógiai Értékelési Konferencia. Szeged, 2012. április 11-13. 82.
- Molnár Gyöngyvér és Józsa Krisztián (2006): Az olvasási képesség értékelésének tesztelméleti megközelítései. In: Józsa Krisztián (szerk.): *Az olvasási képesség fejlődése és fejlesztése*. Dinasztia Tankönyvkiadó, Budapest. 155–174.
- Molnár Gyöngyvér és Kárpáti Andrea (2012): Informatikai műveltség. In: Csapó Benő (szerk.): *Mérlegen a magyar iskola*. Nemzeti Tankönyvkiadó, Budapest. 441–476.
- Molnár Gyöngyvér és Magyar Andrea (2015): A számítógép-alapú tesztelés elfogadottsága pedagógusok és diákok körében. *Magyar Pedagógia*, **115**. 1. sz. 49–66.
- Molnár Gyöngyvér és Pásztor-Kovács Anita (2015a): A számítógépes vizsgáztatás infrastrukturális kérdései: az iskolák eszközparkjának helyzete és a változás tendenciái. *Iskolakultúra*, **25**. 4. sz. 49–61.
- Molnár Gyöngyvér és Pásztor-Kovács Anita (2015b): A problémamegoldó képesség mérése online tesztkörnyezetben. In: Csapó Benő és Zsolnai Anikó (szerk.): *Online diagnosztikus mérések az iskola kezdő szakaszában*. Oktatókutató és Fejlesztő Intézet, Budapest. 279–300.
- Molnár Gyöngyvér, Nikolov Marianne és Csapó Benő (2005): Constructing reading competency scales for second language assessment: Using data of different age groups for IRT scaling. 11th Biennial Conference of EARLI 2005, Nicosia, Cyprus, August 23-27, 2005. 549. o.
- Molnár Gyöngyvér, Papp Zoltán, Makay Géza és Ancsin Gábor (2015): *eDia 2.3 Online mérési platform – feladatfelviteli kézikönyv*. SZTE Oktatáselméleti Kutatócsoport, Szeged.
- Molnár Gyöngyvér, Tongori Ágota és Pluhár Zsuzsa (2015): Az informatikai műveltség online mérése. In: Csapó Benő és Zsolnai Anikó (szerk.): *Online diagnosztikus mérések az iskola kezdő szakaszában*. Oktatókutató és Fejlesztő Intézet, Budapest. 241–260.
- Molnár Pál (2009): Számítógéppel támogatott együttműködő tanulás online közösségi hálózatos környezetben. *Magyar Pedagógia*, 3. sz. 261–285.
- Molnár Pál és Kárpáti Andrea (2009): Az együttműködő tanulás támogatása az oktatási informatika eszközeivel: MapIt vitatérkép. *Új Pedagógiai Szemle*, **59**. 2. sz. 48–60.
- Molnár, G. és Lőrincz, A. (2012): Innovative assessment technologies: Comparing ‘face-to-face’ and game-based development of thinking skills in classroom settings In: Chen, D. (szerk.): *International proceedings of economics development and*

- research. *Management and education innovation*. Vol. 37. IACSIT Press, Singapore. 150–154.
- Molnár, G., Greiff, S. és Csapó, B. (2013): Inductive reasoning, domain specific and complex problem solving: relations and development. *Thinking skills and Creativity*, **9**. 8. sz. 35–45.
- Molnár, G., Greiff, S., Wüstenberg, S. és Fischer, A. (2016): Empirical study of computer based assessment of domain-general dynamic problem solving skills. In: Csapó, B., Funke, J. és Schleicher, A. (szerk.): *On the nature of problem solving*. OECD, Paris. 123–143.
- Moon, J. L. (2013): Comparability of online and paper/pencil mathematics performance measures. PhD disszertáció, College of Education and Human Services, University of Nebraska-Lincoln.
- Moray, N., Lootsteen, P. és Pajak, J. (1986): Acquisition of process control skills. *Systems, Man and Cybernetics, IEEE Transactions*, **16**. 4. sz. 497–504.
- Moss, J. és Hendry, G. (2002): Use of electronic surveys in course evaluation. *British Journal of Educational Technology*, **33**. 5. sz. 583–592.
- Muthén, L. K. és Muthén, B. O. (2010): *Mplus user's guide*. Sixth edition. CA: Muthén & Muthén, Los Angeles.
- Muthén, L. K. és Muthén, B. O. (2012): *Mplus User's Guide* (7th edition). CA: Muthén & Muthén, Los Angeles.
- NAGB (2013): *Technology and engineering literacy framework for the 2014 National Assessment of Educational Progress*. National Assessment Governing Board, WestEd. [http://www.nagb.org/publications/frameworks/prepub\\_naep\\_tel\\_framework\\_2014.pdf](http://www.nagb.org/publications/frameworks/prepub_naep_tel_framework_2014.pdf)
- Nagy József (1971): *Az elemi számolási készségek*. Tankönyvkiadó, Budapest.
- Nagy József (1972): *A témazáró tudásszintmérés gyakorlati kérdései*. Tankönyvkiadó, Budapest.
- Nagy József (1973): *Alapműveleti számolási készségek*. Acta Universitatis Series Specifica, Szeged.
- Nagy József (1975): *A témazáró tesztek reliabilitása és validitása (Standardizált Témazáró Tesztek 18. kötet)*. Acta Universitatis Series Specifica, Szeged.
- Nagy József (1976): *Alsó tagozatos szöveges feladatbank*. JATE, Szeged.
- Nagy József (1987): *A rendszerezési képesség kialakulása*. Gondolkodási műveletek. Akadémiai Kiadó, Budapest.
- Nagy József (1997): Az alapműveltségi vizsga rendszere és általános követelményei. *Új pedagógiai szemle*, **47**. 5. sz. 4–16.
- Nagy József (2000): *XXI. század és nevelés*. Osiris Kiadó, Budapest.
- Nagy József (2007): *Kompetenciaalapú kritériumorientált pedagógia*. Mozaik Kiadó, Szeged.
- Nagy József és Gubán Gyula (1987): A rendszerezési képesség kialakulása és fejlesztése. *Pedagógiai Szemle*, 11. sz. 1108–1118.
- Nagy József, Józsa Krisztián, Vidákovich Tibor és Fazekasné Fenyvesi Margit (2004a). *Diagnosztikus fejlődésvizsgáló és kritériumorientált fejlesztő rendszer 4-8 évesek számára: DIFER programcsomag*. Mozaik Kiadó, Szeged.
- Nagy József, Józsa Krisztián, Vidákovich Tibor és Fazekasné Fenyvesi Margit (2004b): *Az elemi alapképességek fejlődése 4-8 éves életkorban. Az eredményes iskolakezdés*

*hét kritikus alapkészségének országos helyzetképe és a pedagógiai tanulságok.*  
Mozaik Kiadó, Szeged.

- Negroponte, N. (1995): *Being digital*. Vintage, New York.
- Newell, A. és Simon, H. A. (1972): *Human problem solving*. NJ: Prentice-Hall, Englewood Cliffs.
- Neo, M. és Neo, K. T. K. (2001): Innovative teaching: Using multimedia in a problem-based learning environment. *Educational Technology & Society*, **4**. 4. sz. 19–31.
- Novick, L. R. és Bassok, M. (2005): Problem solving. In: Holyoak, K. J. és Morrison, R. G. (szerk.): *The Cambridge handbook of thinking and reasoning*. University Press, Cambridge, NY. 321–349.
- Novick, L. R., Hurley, S. M. és Francis, M. (1999): Evidence for abstract, schematic knowledge of three spatial diagram representation. *Memory and Cognition*, **27**. 288–308.
- Noyes, J. M. és Garland, K. J. (2008): Computer- vs. paper-based tasks: Are they equivalent? *Ergonomics*, **51**. 9. sz. 1352–1375.
- Obama, B. (2016): Giving every student an opportunity to learn through computer science for all. <https://www.whitehouse.gov/photos-and-video/video/2016/01/29/weekly-address-giving-every-student-opportunity-learn-through-comp>
- OECD (2003): *The PISA 2003 assessment framework. Mathematics, reading, science and problem solving knowledge and skills*. OECD, Paris.
- OECD (2004): *Problem solving for tomorrow's World. First measures of cross-curricular competencies from PISA 2003*. OECD, Paris.
- OECD (2006): *Are students ready for a technology rich world? What PISA studies tell us*. OECD, Paris
- OECD (2009a): *PISA CBAS analysis and results – Science performance on paper and pencil and electronic tests*. OECD, Paris.
- OECD (2009b): *PISA 2006 technical report*. OECD, Paris.
- OECD (2010a): *Are the new millennium learners making the grade? Technology use and educational performance in PISA 2006*. <http://www.oecd.org/edu/ceri/45053490.pdf>
- OECD (2010b): *PISA 2009 results: What students know and can do – Student performance in reading, mathematics and science*. <http://www.oecd.org/dataoecd/10/61/48852548.pdf>
- OECD (2010c): *PISA 2012 Field Trial Problem Solving Framework*. <http://www.oecd.org/dataoecd/8/42/46962005.pdf>
- OECD (2011): *PISA 2009 results: Students on line: Digital technologies and performance (Volume VI)*. OECD Publishing, Paris.
- OECD (2012): *Connected minds: Technology and today's learners, educational research and innovation*, OECD Publishing, doi: [10.1787/9789264111011-en](https://doi.org/10.1787/9789264111011-en).
- OECD (2013a): *PISA 2012 Results: What makes schools successful? Resources, policies and practices (Volume IV)*. OECD, Paris.
- OECD (2013b): *PISA 2012 Results. What 15-year-olds know and what they can do with what they know (Volume I)*. OECD, Paris.
- OECD (2013c): *PISA 2012 Assessment and analytical framework: mathematics, reading, Science, problem solving and financial literacy*. OECD Paris.
- OECD (2013d): *PISA 2015 draft collaborative problem solving framework*. OECD, Paris.

- OECD (2013e): *Literacy, numeracy and problem solving in technology-rich environments – Framework for the OECD survey of adult skills*. OECD, Paris.
- OECD (2013f): *OECD skills outlook 2013: First results from the survey of adult skills*. OECD, Paris.
- OECD (2013g): *Key Facts for Hungary in Education at a Glance 2013*. OECD, Párizs. [http://www.oecd.org/edu/Hungary\\_EAG2013%20Country%20Note.pdf](http://www.oecd.org/edu/Hungary_EAG2013%20Country%20Note.pdf)
- OECD (2013h): *Draft mathematics framework*. OECD publishing. <http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Mathematics%20Framework%20.pdf>.
- OECD (2013i): *Draft reading literacy framework*. OECD publishing. <http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Reading%20Framework%20.pdf>
- OECD (2014): *PISA 2012 results: Creative problem solving. Students' skills in tackling real-life problems (Volume V)*. OECD, Paris.
- O'Malley, C. E. (1995, szerk.): *Computer-supported collaborative learning*. Springer-Verlag, Berlin.
- Oregon Department of Education (2007): *Comparability of student scores obtained from paper and computer administrations*. <http://www.ode.state.or.us/teachlearn/testing/manuals/2007/doc4.1comparabilitytesatopandp.pdf>.
- Pachler, N., Daly, C., Mor, Y. és Mellar, H. (2010): *Formative e-assessment: Practitioner cases. Computers and Education*, **54**. 3. sz. 715–721.
- Partnership for 21st Century Skills (2008): *Overview. Framework for 21st century learning. Partnership for 21st century skills*. <http://www.p21.org/overview>.
- Pásztor-Kovács Anita (2016): *A kollaboratív problémamegoldó képesség mérésének elméleti és módszertani megfontolásai: egy pilot kutatás eredményei. Magyar Pedagógia*, **116**. 1. sz. 51–72.
- Pásztor Attila (2014): *Lehetőségek és kihívások a digitális játék alapú tanulásban: egy induktív gondolkodást fejlesztő tréning hatásvizsgálata. Magyar Pedagógia*, **114**. 4. sz. 281–301.
- Pásztor-Kovács Anita, Magyar Andrea, Hülber László, Pásztor Attila és Tongori Ágota (2013): *Áttérés online tesztelésre - a mérés-értékelés új dimenziói. Iskolakultúra*, **23**. 11. sz. 86–100.
- Peak, P. (2005): *Recent trends in comparability studies*. Pearson educational measurement. [http://www.pearsonassessments.com/NR/rdonlyres/5FC04F5A-E79D-45FE-8484-07AACAE2DA75/0/TrendsCompStudies\\_rr0505.pdf](http://www.pearsonassessments.com/NR/rdonlyres/5FC04F5A-E79D-45FE-8484-07AACAE2DA75/0/TrendsCompStudies_rr0505.pdf).
- Pearson (2012): *From paper and pencil to computer-based testing (CBT)*. [http://www.pearsonvue.co.uk/india/Documents/PP\\_to\\_CBT.pdf](http://www.pearsonvue.co.uk/india/Documents/PP_to_CBT.pdf).
- Pearson Educational Measurement (2003): *Virginia standards of learning web-based assessments comparability study report – Spring 2002 administration: Online & paper tests*. Austin.
- Pearson Educational Measurement (2013): *Considerations for next-generation Assessments: A ROADMAP TO 2014*. Austin. [http://www.pearsonassessments.com/hai/images/nextgen/Downloads/NextGen\\_Roadmap\\_Final\\_web.pdf](http://www.pearsonassessments.com/hai/images/nextgen/Downloads/NextGen_Roadmap_Final_web.pdf).

- Pelgrum, W. J. (2004): What can international assessments contribute to help fight low achievement? In: Kárpáti, A. (szerk.): *Promoting equity through ICT in Education*. OECD, Paris, Budapest. 56–69.
- Pellegrino, J. W. és Glaser, R. (1982): Analyzing aptitudes for learning: inductive reasoning. In: Glaser, R. (szerk.): *Advances in instructional psychology*, Vol. 2. NJ: Erlbaum, Hillsdale. 269–345.
- Plichart, P., Jadoul, R., Vandenabeele, L. és Latour, T. (2004): TAO, a Collective distributed computer-based assessment framework built on semantic web standards. In: *Proceedings of the International Conference on Advances in Intelligent Systems – Theory and Application AISTA2004*. IEEE Computer Society, November 15-18, 2004. Luxembourg, Luxembourg.
- Poggio, J. és McJunkin, L. (2012): History, current practice, perspectives, and predictions for the future of computer-based assessment in K-12 education. In: Lissitz, R. W. és Jiao, H. (szerk.): *Computers and their impact on state assessment: Recent history and predictions for the future*. Information Age, Charlotte. 25–53.
- Poggio, J., Glasnapp, D. R., Yang, X. és Poggio, A. J. (2005): A comparative evaluation of score results from computerized and paper and pencil mathematics testing in a large scale stat assessment program. *Journal of Technology, Learning and Assessment*, 3. 6. sz. 4–29.
- Pólya György (1945/1957): *How to solve it*. Princeton University Press, Princeton.
- Pólya György (1969): *A gondolkodás iskolája*. Gondolat Kiadó, Budapest.
- Polya, G. (1954): *Mathematics and plausible reasoning*. Princetown University Press, Princetown.
- Pommerich, M. (2004): Developing computerized versions of paper-and-pencil tests: Mode effects for passage-based tests. *Journal of Technology, Learning, and Assessment*, 2. 6. sz. 1–44.
- Pomplun, M., Ritchie, T. és Custer, M. (2006): Factors in paper-and-pencil and computer reading score differences at the primary grades. *Educational Assessment*, 11. 2. sz. 127–143.
- Popper, K. (1999): *All life is problem solving*. Routledge, London, New York.
- Powers, D. E. és Potenza, M. T. (1996): *Comparability of testing using laptop and desktop computers (RR-96-15)*. Educational Testing Service, Princeton, NJ. [http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content\\_storage\\_01/0000019b/80/14/dc/69.pdf](http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/14/dc/69.pdf). Letöltés ideje: 2009. október 5.
- Putz-Osterloh, W. (1981): Über die Beziehung zwischen Testintelligenz und Problemlöseerfolg. *Zeitschrift für Psychologie*, 189. 79–100.
- Putz-Osterloh, W. és Lüer, G. (1981): Über die Vorhersagbarkeit komplexer Problemlöseleistungen durch Ergebnisse in einem Intelligenztest. *Zeitschrift für experimentellen und angewandte Psychologie*, 28. 309–334.
- Quellmalz, E. S., Timms, M. J., Buckley, B. C., Davenport, J., Loveland, M. és Silberglitt, M. D. (2012): 21st Century dynamic assessment. In: Mayrath, M. C., Clarke-Midura, J., Robinson, D. H. és Schraw, G. (szerk.): *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research*. Information Age Publishing, Charlotte, NC. 55–90.

- Rigas, G., Carling, E. és Brehmer, B. (2002): Reliability and validity of performance measures in microworlds. *Intelligence*, **30**. 463–480.
- R. Tóth Krisztina (2009): A számítógépes és papír-ceruza tesztek eredményeinek összehasonlító vizsgálata háttérváltozók alapján. IX. Országos Neveléstudományi Konferencia, Veszprém, 2009. november 19-21. 97. o.
- R. Tóth Krisztina (2015): Felső tagozatos diákok papír-ceruza és számítógép alapú teszteredményeink összehasonlító vizsgálata szövegértés, induktív gondolkodás és problémamegoldás terén. PhD disszertáció, SZTE Neveléstudományi Doktori Iskola.
- R. Tóth Krisztina és Hódi Ágnes (2011): Számítógépes és papír-ceruza teszteredmények összehasonlító vizsgálata az olvasásszövegértés területén. *Magyar Pedagógia*, **111**. 4. sz. 313–332.
- R. Tóth Krisztina és Molnár Gyöngyvér (2009): A jövő tanárainak IKT ismeretei és elvárásai. *Pedagógusképzés*, **7**. 1. sz. 25–41.
- R. Tóth Krisztina, Molnár Gyöngyvér, Thibaud Latour és Csapó Benő (2011): Az online tesztelés lehetőségei és a TAO platform alkalmazása. *Új Pedagógiai Szemle*, **61**. 1-2-3-4-5. sz. 8–22.
- R. Tóth, K., Molnár, G., Wüstenberg, S., Greiff, S. és Csapó, B. (2011): Measuring adults' dynamic problem solving competency. Paper presented at the 14th European Conference for the Research on Learning and Instruction. Exeter, United Kingdom, August 30- September 3, 2011. In: *Book of abstracts and extended summaries*. 1421–1422.
- Raju, N. S., Laffitte, L. J. és Byrne, B. M. (2002): Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, **87**. 3. sz. 517–529.
- Rasch, G. (1960): *Probabilistic models for some intelligence and attainment tests*. Danish Institute for Educational Research, Copenhagen.
- Raven, J. C. (1962): *Advanced Progressive Matrices, Set II*. H. K. Lewis, London.
- Redecker, C. és Johannessen, Ø. (2013): Changing assessment – towards a new assessment paradigm using ICT. *European Journal of Education*, **1**. 48. sz. 79–96.
- Reeff, J. P., Zabal, A. és Blech, C. (2006): *The assessment of problem-solving competencies*. Deutsches Institut für Erwachsenenbildung, Bonn. [http://www.die-bonn.de/espid/dokumente/doc-2006/reeff06\\_01.pdf](http://www.die-bonn.de/espid/dokumente/doc-2006/reeff06_01.pdf).
- Revákné Markóczi Ibolya (2001): A problémamegoldó gondolkodást befolyásoló tényezők. *Magyar Pedagógia*, **101**. 3. sz. 267–285.
- Revákné Markóczi Ibolya, Tóth Zoltán és Tóthné Kosztin Beáta (2009): Természettudományos problémamegoldó stratégiák vizsgálata az általános iskola alsó tagozatában. IX. Országos Neveléstudományi Konferencia, Veszprém 2009. november 19-21.
- Ridgway, J. (2003): Standard Setting via the Fat Anchor. <https://www.dur.ac.uk/resources/smart.centre/Publications/jrIAEAfatanchorfinalpaper.doc>.
- Ridgway, J. és McCusker, S. (2003): Using computers to assess new educational goals. *Assessment in Education*, **10**. 3. sz. 309–328.
- Ripley, M. (2010): Transformational Computer-based Testing. In Scheuermann, F. és Björnsson, J. (szerk.): *The transition to computer-based assessment: New*

- approaches to skills assessment and implications for large-scale testing*. European Communities, Brussels. 92–98.
- Rose, M., Hess, V., Hörhold, M., Brähler, E. és Klapp, B. F. (1999): Mobile computergestützte psychometrische Diagnostik. Ökonomische Vorteile und Ergebnisse zur Teststabilität. *Psychotherapie Psychosomatik Medizinische Psychologie*, **49**. 202–207.
- Rost, J. (2004): *Lehrbuch Testtheorie – Testkonstruktion*. Verlag Hans Huber. Bern, Switzerland.
- RTTT (2009): Race to the Top Program Executive Summary. <https://www2.ed.gov/programs/racetothetop/executive-summary.pdf>.
- Sandene, B., Bennett, R. E., Braswell, J. és Oranje, A. (2005): Online assessment in mathematics. In: Sandene, B., Horkay, N., Bennett, R. E., Allen, N., Braswell, J., Kaplan B. és Oranje, A. (szerk.): *Online assessment in mathematics and writing: Reports from the NAEP technology-based assessment project (NCES 2005-457)*. National Center for Education Statistics, US Department of Education, Washington, DC. 1–67.
- Sandene, B., Horkay, N., Bennett, R., Allen, N., Braswell, J., Kaplan, B. és Oranje, A. (2005): *Online Assessment in Mathematics and Writing: Reports From the NAEP Technology-Based Assessment Project, Research and Development Series (NCES 2005-457)*. U.S. Department of Education, National Center for Education Statistics, U.S. Government Printing Office, Washington, DC.
- Scardamalia, M. (2002): Collective cognitive responsibility for the advancement of knowledge. In: Smith, B. (szerk.): *Liberal education in a knowledge society*. IL: Open Court, Chicago. 67–98.
- Scheuermann, F. és Björnsson, J. (2009, szerk.): *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing*. Office for Official Publications of the European Communities, Luxembourg.
- Scheuermann, F. és Pedró, F. (2009, szerk.): *Assessing the effects of ICT in education. Indicators, criteria and benchmarks for international comparisons*. Office for Official Publications of the European Communities, Luxembourg.
- Scheuermann, F. és Pereira, G. A. (2008, szerk.): *Towards a research agenda on computer-based assessment*. Office for Official Publications of the European Communities, Luxembourg.
- Snow, K. L. (2006): Measuring school readiness: Conceptual and practical considerations. *Early Education and Development*, **17**. 1. sz. 7–41.
- Snow, C. E. és Van Hemel, S. B. (2008, szerk.): *Early childhood assessment: Why, what, and how*. The National Academies Press, Washington, DC.
- Schraw, G., Dunkle, M. E. és Bendixen, L. D. (1995): Cognitive processes in well-defined and ill-defined problem solving. *Applied Cognitive Psychology*, **9**. 6. sz. 523–538.
- Schroeders, U. és Wilhelm, O. (2011): Equivalence of reading and listening comprehension across test media. *Educational and Psychological Measurement*, **71**. 5. sz. 849–869.
- Scottish Qualifications Authority (2003): *Key competencies — some international comparisons*. Scottish Qualifications Authority, Glasgow. [http://www.sqa.org.uk/files\\_ccc/Key\\_Competencies.pdf](http://www.sqa.org.uk/files_ccc/Key_Competencies.pdf)



- Selwyn, N. (2011): *Education and technology. Key issues and debates*. A&C Black, Chennai, India.
- SETDA (2008): SETDA's national trends report 2008. Enhancing education through technology. <http://www.setda.org/wp-content/uploads/2013/12/National-Trends-Report-2008.pdf>.
- Shaw, S., Crisp, V. és Johnson, N. (2012): A framework for evidencing assessment validity in large-scale, high-stakes international examinations. *Assessment in Education: Principles, Policy & Practice*, **2**. 19. sz. 159–176.
- Sim, G. és Horton, M. (2005): Performance and attitude of children in computer based versus paper based testing. In: Kommers, P. és Richards, G. (szerk.): *Proceedings of world conference on educational multimedia, hypermedia and telecommunications 2005*. VA: AACE, Chesapeake. 3610–3614.
- Simon, H. (1974): Problem solving and rule induction: a unified view. In: Gregg, L.W. (szerk.): *Knowledge and cognition*. MA: Erlbaum, Potomac. 105–128.
- Sonnleitner, P., Brunner, M., Greiff, S., Funke, J., Keller, U., Martin, R., Hazotte, C., Mayer, H. és Latour, T. (2012): The Genetics Lab. Acceptance and psychometric characteristics of a computer-based microworld to assess complex problem solving. *Psychological Test and Assessment Modeling*, **54**. 54–72.
- Sørensen, H. és Andersen, A. M (2009): How did Danish students solve the PISA CBAS items? Right and wrong answers from a gender perspective. In: Scheuermann, F. és Björnsson, J. (szerk.): *The transition to computer-based assessment. New approaches to skills assessment and implications for large-scale testing*. Office for Official Publications of the European Communities, Luxembourg. 194–207.
- Steenkamp, J.-B. E. M. és Baumgartner, H. (1998): Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, **25**. 1. sz. 78–90.
- Steinmetz, H., Schmidt, P., Tina-Booh, A., Wieczorek, S. és Schwartz, S. H. (2007): Testing measurement invariance using multigroup CFA: Differences between educational groups in human values measurement. *Quality and Quantity*, **43**. 4. sz. 599– 616.
- Steklács János, Molnár Gyöngyvér és Csapó Benő (2015): Az olvasás-szövegértés online diagnosztikus mérések tartalmi kereteinek elméleti háttere. In: Csapó Benő, Steklács János és Molnár Gyöngyvér (szerk.): *Az olvasás-szövegértés online diagnosztikus értékelésének tartalmi keretei*. Oktatókutató és Fejlesztő Intézet, Budapest. 13–28.
- Sternberg, R. (1994, szerk.): *Thinking and problem solving*. Academic Press, San Diego.
- Strain-Seymour, E., Way, W. D. és Dolan, R. P. (2009): *Strategies and processes for developing innovative items in large-scale assessments*. Pearson Education, Iowa City, IA.
- Suhr, K. A., Hernandez, D. A., Grimes, D. és Warschauer, M. (2010): Laptops and fourth Grade literacy: Assisting the jump over the fourth-Grade slump. *The Journal of Technology, Learning, and Assessment*, **9**. 5. sz. <http://escholarship.bc.edu/jtla/vol9/5/>
- Szabó Gábor (2008): *Applying item response theory in language test item bank building*. Peter Lang, München.
- Taylor, C., Kirsch, I., Eignor, D. és Jamieson, J. (1999): Examining the relationship between computer familiarity and performance on computer-based language tasks. *Language Learning*, **49**. 219–274.

- Terzis, V. és Economides, A. A. (2011): The acceptance and use of computer based assessment. *Computers & Education*, **4**. 56. sz. 1032–1044.
- Texas Education Agency (2008): *A review of literature on the comparability of scores obtained from examinees on computer-based and paper-based tests*. Texas Education Agency (TEA) Technical Report Series, Texas.
- Thompson, N. és Wiess, D. (2009): Computerised and a daptive testing in educational assessment. In: Sheuermann, F. és Björnsson, J. (szerk.): *The transition to computer-based assessment. New approaches to skills assessment and implications for large-scale testing*. Office for Official Publications of the European Communities, Luxemburg. 127–133.
- Tongori Ágota (2012): Az IKT műveltség fogalmi keretének változása. *Iskolakultúra*, **22**. 11. sz. 34–47.
- Tóth Edit (2009): A tanárok értékeléssel kapcsolatos attitűdjei, nézetei. In: Bárdos Jenő és Sebestyén József (szerk.): *IX. Országos Neveléstudományi Konferencia: Neveléstudomány - Integritás és integrálhatóság. Program és tartalmi összefoglalók*. Magyar Tudományos Akadémia, Pedagógiai Bizottság, Veszprém. 206–207.
- Tóth Edit (2010): Teszt alapú elszámoltathatóság a közoktatásban. *Iskolakultúra*, **20**. 1. sz. 60–79.
- Tóth Edit (2011): Pedagógusok nézetei a tanulóitелjesítmény-mérésekről. *Magyar Pedagógia*, **111**. 3. sz. 225–249.
- Tóth Edit, Molnár Gyöngyvér és Csapó Benő (2011): Az iskolák IKT felszereltsége – helyzetkép országos reprezentatív minta alapján. *Iskolakultúra*, **21**. 10–11. 124–137.
- Tóth Krisztina és Hódi Ágnes (2013): A mérőeszköz-bővítéstől a tesztelési folyamat vizsgálataig: számítógépes tesztelés nagymintás nemzetközi vizsgálatokban. *Iskolakultúra*, **23**. 9. sz. 75–88.
- Tóth, K., Rölke, H. és Goldhammer, F. (2012): Educational process mining. Clustering students' test-taking behaviour in internet-based simulations. X. Pedagógiai Értékelési Konferencia, Szeged, 2012. április 26-28. 44.
- Tóth K., Rölke, H., Greiff, S. és Wüstenberg, S. (2014): Discovering students' complex problem solving strategies in educational assessment. In: Stamper, J., Pardos, Z., Mavrikis, M. és McLaren, B. M. (szerk.): *Proceedings of the 7th International Conference on Educational Data Mining*. International Educational Data Mining Society. 225–228
- Trilling B. és Fadel. C. (2009): *21st Century skills. Learning for life in our time*. Jossey-Bass A Wiley Imprint. San Francisco.
- Tschirgi, J. E. (1980): Sensible reasoning: A hypothesis about hypotheses. *Child Development*, **51**. 1–10.
- Ullman, J. B. (2007): Structural equation modeling. In: Tabachnick, B.G. és Fidell, L. S. (szerk.): *Using multivariate statistics*. Allyn Bacon, New York. 653–771.
- UNESCO (2002): *Information and communication technologies in teacher education*. UNESCO, Paris.
- UNESCO (2008): *Towards Information Literacy indicators*. UNESCO, Paris.

- UNESCO (2014): Information and communication technology (ICT) in education in Asia. A comparative analysis of ICT integration and e-readiness in schools across Asia. UNESCO, <http://www.uis.unesco.org/Communication/Documents/ICT-asia-en.pdf>.
- UNESCO és Microsoft (2011): *Unesco ICT Competency Framework for Teachers*. UNESCO, Paris.
- Valenti, S., Neri, F. és Cucchiarelli, A. (2003): An overview of current research on automated essay grading. *Journal of Information Technology Education: Research*, **2**. 1. sz. 319–330.
- Van der Kleij, F. M., Eggen, T. J. H. M., Timmers, C. F., és Veldkamp, B. P. (2012): Effects of feedback in a computer-based assessment for learning. *Computers & Education*, **58**. 1. sz. 263–272.
- van Lent, G. (2010): Risks and benefits of CBT versus PBT in high-stakes testing. In: Scheuermann, F. és Bjornsson, J. (szerk.): *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing*. European Communities, Brussels. 83–91.
- Vandenberg, R. J. és Lance, C. E. (2000): A review and synthesis of the MI literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, **3**. 1. sz. 4–70.
- Vári Péter (1997): A tanulói teljesítmények mérése itthon és külföldön. In: Vári Péter (szerk.): *MONITOR '95 – A tanulók tudásának felmérése*. OKI, Budapest. 13–42.
- Vidákovich Tibor (1987): *Innovatív célú diagnosztikus pedagógiai értékelés*. Közoktatási Kutatások Titkársága, Budapest.
- Vidákovich Tibor (1990): *Diagnosztikus pedagógiai értékelés*. Akadémiai Kiadó, Budapest.
- Vígh Tibor (2008): Egy IRT-alapú nyelvi feladatbank létrehozásának módszertani kérdései. A német nyelvi érettségi vizsgafeladatok eredményei. *Magyar Pedagógia*, **108**. 1. sz. 29–51.
- Wallace, P. E. és Clariana, R. B. (2005): Test mode familiarity and performance-gender and race comparisons of test scores among computer-literate students in advanced information systems courses. *Journal of Information Systems Education*, **16**. 2. sz. 177–182.
- Wang, S. (2004): *Online or paper: does delivery affect results? Administration mode comparability study for Stanford Diagnostic Reading and Mathematic Tests*. Harcourt Assessment Inc, USA.
- Wang, S., Jiao, H., Young, M., Brooks, T. és Olson, J. (2008): Comparability of computer-based and paper-and-pencil testing in K-12 reading assessments: A meta-analysis of testing mode effects. *Educational and Psychological Measurement*, **68**. 1. sz. 5–24.
- Wastiau, P., Blamire, R., Kearney, C., Quittre, V., Van der Gaer, R. és Monseur, C. (2013): The use of ICT in education: a survey of schools in Europe. *European Journal of Education*, **48**. 1. sz. 11–27.
- Waters, S. D. és Pommerich, M. (2007): Context effects in internet testing: A literature review. Előadás. 22nd Annual Conference of the Society for Industrial and Organizational Psychology. 2007. április 7. New York City, Amerikai Egyesült Államok.

- Way, W. D., Davis, L. L. és Fitzpatrick, S. (2006): Score comparability of online and paper administrations of Texas assessment of knowledge and skills. Előadás. Annual meeting of National Council on Measurement in Education, 2006. április, San Francisco, Amerikai Egyesült Államok.
- Wenke, D., Frensch, P. A. és Funke, J. (2005): Complex problem solving and intelligence. In: Sternberg, R. J. és Pretz, J. E. (szerk.): *Cognition and intelligence*. Cambridge University Press, Cambridge. 160–187.
- Wertheimer, M. (1945): *Productive Thinking*. Harper, New York, London.
- White, G. (2008): *ICT trends in education. Teaching and learning and leadership digital learning research*. Australian Council for Educational Research. [http://research.acer.edu.au/cgi/viewcontent.cgi?article=1002&context=digital\\_learning](http://research.acer.edu.au/cgi/viewcontent.cgi?article=1002&context=digital_learning).
- Williamson, D. M., Mislevy, R. J. és Bejar, I. I. (2006, szerk.): *Automated scoring of complex tasks in computer-based testing*. Erlbaum, Mahwah, NJ.
- Wilton, J. A. és McLean, R. S. (1984): Evaluation of a mouse as an educational pointing device. *Computers & Education*, **8**. 4. sz. 455–461.
- Winter, P. C. (2010): *Evaluating the comparability of scores from achievement test variations*. Council of Chief State School Officers. Amerikai Egyesült Államok. [http://www.ccsso.org/Documents/2010/Evaluating\\_the\\_Comparability\\_of\\_Scores\\_2010.pdf](http://www.ccsso.org/Documents/2010/Evaluating_the_Comparability_of_Scores_2010.pdf).
- Wirth, J. és Klieme, E. (2004): Computer-based assessment of problem solving competence. *Assessment in Education: Principles, policy and Practice*, **10**. 3. sz. 329–345.
- Wise, S. L. és Plake, B. S. (1990): Computer-based testing in higher education. *Measurement and Evaluation in Counseling and Development*, **23**. 3–10.
- World Economic Forum (2010): *The Lisbon review 2010*. World Economic Forum, Switzerland. <http://www.weforum.org/pdf/Gcr/LisbonReview/TheLisbonReview2010.pdf>.
- Write, B. D. és Masters, G. N. (1982): *Rating scale analysis*. MESA press, Chicago.
- Write, B. D. és Stone, M. H. (1979): *Best test design*. MESA press, Chicago.
- Wu, M. (2006): *PISA training workshop: Application of item response theory (IRT) to PISA (ConQuest)*. Hong Kong PISA Centre, Hong Kong.
- Wu, A. D., Li, Z. és Zumbo, B. D. (2007): Decoding the meaning of factor invariance and updating the practice of multi-group confirmatory factor analysis. *Practical Assessment, Research and Evaluation*, **12**. 3. sz. 1–26.
- Wüstenberg, S., Greiff, S. és Funke, J. (2012): Complex problem solving — More than reasoning? *Intelligence*, **40**. 1–14.
- Wüstenberg, S., Greiff, S., Molnár, G. és Funke, J. (2014): Determinants of cross-national gender differences in complex problem solving competency. *Learning and Individual Differences*, **29**. 18–29.
- Wüstenberg, S., Stadler, M., Hautamäki, J. és Greiff, S. (2014): The role of strategy knowledge for the application of strategies in complex problem solving tasks. *Technology, Knowledge, and Learning*, **19**. 127–146.
- Yin, L., Chen, X., Sun, Y. Worm, T. és Reale, M. (2008): A high-resolution 3D dynamic facial expression database. In: 8th IEEE International Conference on Automatic Face Gesture Recognition, IEEE. Number: 66. 1–6.

- Zelman, M., Avdeeva, S., Shmis, T., Vasiliev, K. és Froumin, I. (2011): International Comparison of Information Literacy in Digital Environments. Conference Papers. International Association for Educational Assessment (IAEA) Manila, 37th annual conference 2011. [http://www.iaea.info/documents/paper\\_30e43f54.pdf](http://www.iaea.info/documents/paper_30e43f54.pdf)
- Zhang, Y., Powers, D. E., Wright, W. és Morgan, R. (2003): *Applying the online scoring network (OSN) to advanced placement program (AP) tests (RM-03-12)*. Educational Testing Service, Princeton, NJ. <http://www.ets.org/research/researcher/RR-03-12.html>.
- Zsolnai Anikó és Józsa Krisztián (2002): A szociális készségek kritériumorientált fejlesztésének lehetőségei. *Iskolakultúra*, **12.** 4. sz. 12–20.

## KÖSZÖNET

Mindenek előtt kiemelt köszönettel tartozom *Csapó Benő*nek, aki egyrészt felhívta a figyelmem az empirikus neveléstudományi kutatások jelentőségére, másrészt egész eddigi pályám során folyamatosan biztatott, támogatott és hasznos javaslataival segítette kutatásaimat.

Köszönettel tartozom a disszertációban bemutatott hazai kutatások intézményi keretét biztosító Oktatáselméleti Kutatócsoport, MTA–SZTE Képességfejlődés Kutatócsoport és SZTE Neveléstudományi Intézet munkatársainak. Külön kiemelném *Molnár Katalin*, *Kléner Judit*, *Bogárné Németh Mária*, *Patai Jolán*, *Sisák Bernadett*, *Nagy Zsuzsanna*, *Fűz Nóra* és *Zergényi Barbara* kutatások szervezésében és pályázatok kivitelezésében nyújtott segítségét.

Az eDia online platformfejlesztések kapcsán nélkülözhetetlen kollégáim voltak: *Makay Géza* és *Ancsin Gábor*, míg az adatfelvételek technikai háttérét *Betyár Gábor*, *Halof Ferenc* és *Papp Zoltán* biztosította.

Köszönettel tartozom *Lőrincz András*nak, aki felhívta a figyelmem a kontextuális adatok elemzésében rejlő lehetőségekre és egy közös projekt keretein belül megvalósítottuk egy korábbi szemtől szemben alkalmazott gondolkodásfejlesztő tréning számítógépes játék-alapú megjelenítését.

Köszönöm *Nagy József* támogatását, a hasznos szakmai beszélgetéseket, melyekkel sok tekintetben megerősítette munkám.

A hazai kutatókon, kollégákon kívül köszönettel tartozom számos külföldi kollégának, akik hasznos ötleteikkel, meglátásaikkal, közös projektek megvalósításával támogatták munkámat, segítették fejlődésem: *Joachim Funke*, *Samuel Greiff*, *Trevor Bond*, *Margaret Wu*, *Sascha Wüstenberg*, *Johannes Hartig*, *Andreas Frey*, *Sirkku Kupiainen*, *Jean-Paul Reef*, *Thibaud Latour*.

Köszönettel tartozom tanítványaimnak, akik kutatásaikkal, észrevételeikkel, a közös munkával hozzájárultak fejlődésemhez és a disszertáció egyes részeiben bemutatott eredmények létrejöttéhez: *Tóth Krisztina*, *Hülber László*, *Magyar Andrea*, *Szili Katalin*, *Pásztor-Kovács Anita*, *Pásztor Attila*, *Tóth Alisa*, *Nagy Gyula*, *Wu Hao*.

A disszertációban bemutatott kutatások az alábbi támogatások segítségével valósultak meg: Diagnosztikus mérések fejlesztése első fázis (TÁMOP-3.1.9-08/1-2009-0001) és Diagnosztikus mérések fejlesztése (második fázis, TÁMOP-3.1.9-11/1-2012-0001) kiemelt projekt, OTKA (K75274) A komplex problémamegoldó gondolkodás különböző mérési lehetőségeinek összehasonlító elemzése és OTKA (K111945) Defining and enhancing the component skills of problem solving projektek, továbbá Magyar Tudományos Akadémia Bolyai János Kutatási Ösztöndíj 2005–2008 és 2009–2012 között.

Végül, de nem utolsónak kiemelt köszönettel tartozom a családomnak, férjemnek, három gyermekemnek, akik sok-sok szeretettel és lemondással támogatták kutatói pályámat.

## RÖVIDÍTÉSEK JEGYZÉKE

aBIC	adjusted Bayesian Information Criterion (korrigált Bayesi Információs Kritérium)
AIC	Akaike Information Criterion (Akaike Információs Kritérium)
ATCS21	Assessment and Teaching 21st Century Skills
BIC	Bayesian Information Criterion (Bayesi Információs Kritérium)
CB	Computer-based (számítógép-alapú)
CBA	Computer-based Assessment (számítógép-alapú mérés-értékelés)
CFI	Comparative Fit Index (Komparatív Illeszkedési Mutató)
df	degrees of freedom (szabadsági fok)
DIFFTEST	Difference Test procedure (a modellek különbözőségének tesztje)
DP	dinamikus problémamegoldó
DPK	dinamikus problémamegoldó képesség
eDia	electronic Diagnostic Assessment System
FF	Face-to-face (szemtől szemben)
IEA	International Association for the Evaluation of Education Achievement (Az oktatási teljesítmény értékelésnek nemzetközi szervezete)
IKT	Information and Communication Technology (Információs és kommunikációs technológiák)
IG	Induktív gondolkodás
L-M-R teszt	Lo-Mendell-Rubin Adjusted Likelihood Ratio Test (Lo-Mendell-Rubin valószínűségi hányados teszt)
MGCFA	Multi-Group Confirmatory Factor Analysis (Többcsoportos faktoranalízis)
NAEP	National Assessment of Educational Progress
OECD	Organisation for Economic Co-operation and Development (Gazdasági Együttműködési és Fejlesztési Szervezet)
PIRLS	Progress in International Reading Literacy Study (Nemzetközi szövegértési vizsgálat)
PISA	Programme for International Student Assessment (Nemzetközi tanulói teljesítménymérés programja)
PP	Paper-and-pencil (papíralapú)
PSR	Person Separation Reliability
PV	plausible values
RMSEA	Root mean square error approximation (a megközelítés hibáját a modell komplexitását figyelembe vevő módon mérő index)
sd	Standard Deviation (szórás)
SEM	Structural equation modeling (strukturális egyenlet modellezés)
TAO	Testing Assisté par Ordinateur (számítógép-alapú tesztelő platform)
TBA	Technology-based Assessment (technológiaalapú mérés-értékelés)
TIMSS	Trends in International Mathematics and Science Study (A matematika és természettudományos kutatások trendjei)
TLI	Tucker-Lewis Fit Index (Tucker-Lewis Illeszkedési Index)
WLE	Weighted likelihood estimation

WLSMV	Weighted Least Squares Model Variation (súlyozott legkisebb négyzetes becslési eljárás)
$\Delta\text{CFI}$	Delta Comparative Fit Index (Komparatív Illeszkedési Mutatók különbsége)
$\Delta\chi^2$	Delta Chi-square (A Khí-négyzet próba különbsége)
$\chi^2$	Chi-square test (Khí-négyzet próba)
z	z-test (z-próba, Fisher r-to-z transzformáció eredménye)



## TÁBLÁZATOK JEGYZÉKE

2.1. táblázat.	Kutatások, melyek a nem PP és CB teszteken nyújtott teljesítménybefolyásoló hatását vizsgálták .....	49
2.2. táblázat.	A disszertációban bemutatásra kerülő empirikus kutatások listája és főbb jellemzői .....	59
3.1. táblázat.	A két adatfelvétel mintájának főbb jellemzői .....	63
3.2. táblázat.	A mérési modell illeszkedésmutatóinak jósága, a billentyűzet és egérhasználati képességek dimenzionalitás-vizsgálata.....	67
3.3. táblázat.	A kattintás, a vonszolás és a gépelés műveleteinek időkorláttal és időkorlát nélkül történő kivitelezési sikeressége első évfolyamon .....	70
3.4. táblázat.	A diákok egér és billentyűzethasználati képességszintjének változása évfolyamonkénti bontásban (mle) .....	72
3.5. táblázat.	A DIFER szemtől szembeni és számítógépes adatfelvételének mintája .....	76
3.6. táblázat.	Kategoriális változók esetén a mérési invariancia tesztelésének lépései.....	81
3.7. táblázat.	Az egyéni és a számítógép-alapú adatfelvétel eredményeinek megbízhatósági mutatói .....	82
3.8. táblázat.	A beszédhanghallás (BHH) teszt elméleti modelljeinek illeszkedésmutatói (számítógép alapon).....	83
3.9. táblázat.	A tesztek mérési invarianciáját jellemző modellek illeszkedésmutatói .....	85
3.10. táblázat.	Az egyéni és számítógépes adatfelvétel eredményei közötti különbségek .....	85
3.11. táblázat.	Az induktív gondolkodás teszt mérési invarianciáját jellemző modellek illeszkedésmutatói .....	89
4.1. táblázat.	Az adatelemzésekbe bevont kutatások mintáinak főbb tulajdonságai.....	103
4.2. táblázat.	A dinamikus problémamegoldó képességet mérő tesztekkel történő kutatások mintáinak főbb tulajdonságai .....	118
4.3. táblázat.	Az öt adatfelvétel során alkalmazott szcenárió főbb jellemzői .....	123
4.4. táblázat.	A CFT teszt mérési invariancia-vizsgálatának illeszkedésmutatói .....	124
4.5. táblázat.	Az egyetemista bemeneti mérés során alkalmazott tesztek reliabilitásmutatói az itemkihagyásos elemzések után .....	124
4.6. táblázat.	A DPK tesztváltozatok reliabilitásmutatói .....	126
4.7. táblázat.	A DPK teszt dimenzionalitás-vizsgálatának illeszkedésmutatói.....	128
4.8. táblázat.	A MicroDYN megközelítés fényében felépített DPK teszt mérési invariancia-vizsgálatának illeszkedésmutatói.....	129
4.9. táblázat.	A MicroDYN megközelítés fényében felépített DPK teszt nemek szerinti mérési invariancia-vizsgálatának illeszkedésmutatói .....	130
4.10. táblázat.	Hetedik évfolyamon a négy adatfelvétel közös itemei alapján számolt alapstatisztikai mutatók .....	135
5.1. táblázat.	A statikus és a dinamikus problémahelyzetekben mutatott teljesítmény és az intelligencia kapcsolata évfolyamonkénti bontásban .....	146
5.2. táblázat.	A DPK teszten elért teljesítmények nemek szerinti bontásban .....	151

5.3. táblázat.	A 6-8. évfolyamosok tesztjében szereplő problémák minimális komplexitása .....	166
5.4. táblázat.	Az elméletileg helyes vagy a rendszer kiismeréséhez nem elegendő stratégia alkalmazásának sikeressége a problémák komplexitásának fényében (minimális stratégia) .....	167
5.5. táblázat.	Az elméletileg helyes vagy a rendszer kiismeréséhez nem elegendő stratégia alkalmazásának sikeressége a problémák komplexitásának fényében 6. és 8. évfolyamon (minimális stratégia) .....	168
5.6. táblázat.	Az elméletileg helyes stratégiákon belül a VOTAT, a VOTAT stratégiákon belül az elszigetelt változókezelést alkalmazó VOTAT, valamint a nem VOTAT stratégiák aránya .....	169
5.7. táblázat.	A tudatos és kevésbé tudatos VOTAT-stratégia használók problémamegoldó sikeressége .....	175
5.8. táblázat.	A látens profil elemzések illeszkedési mutatói a 6-tételes stratégiaelemzés esetén (6-8-os minta) .....	176
5.9. táblázat.	Az adott profilhoz tartozás valószínűsége .....	177
5.10. táblázat.	A különböző profillal jellemezhető problémamegoldó csoportokhoz tartozás gyakorisága évfolyamonkénti bontásban.....	178
5.11. táblázat.	A látens profil elemzések illeszkedési mutatói a 6-tételes stratégiaelemzés esetén (7-es longitudinális minta) .....	179
5.12. táblázat.	A látens profil elemzések illeszkedési mutatói a diákok problémamegoldó stratégiahasználatának elemzése alapján (egyetemista minta) .....	181
5.13. táblázat.	A kivitelezett manipulációk száma és a probléma megoldásával töltött idő összefüggése az alkalmazott stratégia helyességével és a teszt egyes problémáin mutatott teljesítménnyel .....	183
5.14. táblázat.	A teszt megoldása során az egymás utáni problémák feltérképezése során alkalmazott manipulációk száma közötti összefüggések.....	184

## ÁBRÁK JEGYZÉKE

1.1. ábra.	Az ipari szektorban történő foglalkoztatás jellegének változása 1980 és 2007 között (forrás: OECD, 2013f. 48. o.).....	8
2.1. ábra.	Az itemkarakterisztikus görbe .....	19
2.2. ábra.	Három teszt horgonyzási lehetőségei (1 teszt 2 résztesztből áll – szűkebb képességtartomány mérése) .....	21
2.3. ábra.	Három teszt horgonyzási lehetőségei (1 teszt 2 résztesztből áll – tág képességtartomány mérése) .....	22
2.4. ábra.	A PISA 2006 mérés során alkalmazott klaszterkapcsolatok (S- természettudomány, M- matematika, R-olvasás; a számok a területen belül alkalmazott klaszter sorszáma) .....	22
2.5. ábra.	A feladatszintű adaptív tesztelés menetének illusztrációja. A pontok a feladatok nehézségi szintjét reprezentálják. ....	30
2.6. ábra.	Példa részteszt szintű adaptív tesztelés megvalósítására (RT: részteszt) .....	32
2.7. ábra.	Az adaptív és a lineáris formátumú teszt standard hibáinak alakulása a tanulók képességszintjének függvényében (Forrás: Magyar, 2014b. 59. o.) ...	32
2.8. ábra.	A személy képességszintje és a mérési pontosság közötti összefüggés adott nehézségű teszt esetén (Jude, 2006 alapján).....	34
2.9. ábra.	A lineáris és az adaptív technikával összeállított, azonos nehézségű modulokat tartalmazó tesztek információs függvényei (Forrás: Magyar és Molnár, 2015) .....	35
2.10. ábra.	Az itemfejlesztés költségei PP: papíralapú, CA: papíralapú, de számítógéppel segített, TM: számítógép-alapú informatikai szakemberek által programozott feladatok, PF: számítógép-alapú, feladatírók által szerkesztett feladatokkal (Forrás: Farcout és Latour, 2009. 113. o.) .....	38
2.11. ábra.	Az itemírás szervezeti költségei (PP: papíralapú, CA: papíralapú, de számítógéppel segített, TM: számítógép-alapú IT szakemberek által programozott feladatok, PF: számítógép-alapú, feladatírók által szerkesztett feladatokkal) (Forrás: Farcout és Latour, 2009. 113. o.) .....	39
2.12. ábra.	Az itemírás biztonsági költségei (PP: papíralapú, CA: papíralapú, de számítógéppel segített, TM: számítógép-alapú IT szakemberek által programozott feladatok, PF: számítógép-alapú, feladatírók által szerkesztett feladatokkal) (Forrás: Farcout és Latour, 2009. 114. o.) .....	39
2.13. ábra.	A technológiaalapú tesztelés trendjei (Redecker és Johannessen, 2013 alapján) .....	50
2.14. ábra.	Az OECD PISA 2012 számítógépes matematika eredmények (forrás: Balázsi, Ostorics, Szalay, Szepesi és Vadász, 2013).....	54
2.15. ábra.	Az OECD PISA 2012 digitális szövegolvasás eredmények (forrás: Balázsi, Ostorics, Szalay, Szepesi és Vadász, 2013).....	54

3.1. ábra.	Egérkezelést mérő példafeladat (képelemeken történő navigálás és kattintás – pontosság) .....	64
3.2. ábra.	Egérkezelést mérő példafeladat (választógombra történő kattintás – pontosság és gyorsaság) .....	64
3.3. ábra.	Egérkezelést mérő példafeladat (vonszolás pontossága).....	65
3.4. ábra.	Billentyűzetkezelést mérő példafeladat (gépelés pontossága).....	66
3.5. ábra.	Az egér- és billentyűzet-használati képességeket mérő feladatok személy-item térképe (minden 'x' 53 diákot reprezentál).....	68
3.6. ábra.	A kattintás, a vonzolás és a gépelés műveletei sikerességének kapcsolata első évfolyamon (korrelációs és parciális korrelációs értékek, az ábrán szereplő minden együttható $p < 0,01$ szinten szignifikáns).....	69
3.7. ábra.	Mintafeladat kattintással történő színezésre .....	71
3.8. ábra.	A kattintás, a vonzolás és a gépelés műveletei nehézségi indexének változása a teszt első és azonos műveletet kívánó második itemén (jelmagyarázat rövidítései: katt: kattintás, dnd: 'drag-and-drop', gép: gépelés, k: kicsi, n: nagy, e: egy elem, t: több elem: k: időkorlátos, kn: időkorlát nélküli).....	71
3.9. ábra.	Mintafeladat a beszédhanghallás tesztből egyéni adatfelvétel és számítógép-alapú adatfelvétel esetén .....	77
3.10. ábra.	Példafeladat a relációsózkincs teszt számítógépes változatából.....	78
3.11. ábra.	Egy-egy példafeladat az elemi számolási készség tesztből .....	78
3.12. ábra.	Mintafeladat az 1. évfolyamosok számára készült tesztből.....	79
3.13. ábra.	A beszédhanghallás teszt egy és négydimenziós elméleti faktorstruktúrája .....	84
3.14. ábra.	A tanulók hagyományos illetve számítógépes módban nyújtott teljesítményei a beszédhanghallás teszten .....	86
3.15. ábra.	Az IRT modellel számított itemparaméterek összehasonlítása a beszédhanghallás teszt egyéni és számítógépes adatfelvételi módjában.....	87
3.16. ábra.	A tapasztalati összefüggés-megértés és a tapasztalati következtetés teszt itemeinek viselkedésváltozása a hagyományos és számítógép-alapú adatfelvétel során .....	88
3.17. ábra.	Az induktív gondolkodás teszt itemeinek nehézségi indexe papír, illetve számítógépes környezetben .....	89
4.1. ábra.	Példafeladat a PISA 2003-as Problémamegoldás moduljából (papíralapú) .....	97
4.2. ábra.	Egy tipikus MicroDYN probléma szerkezete három bemeneti (A, B, C) és három kimeneti (X, Y, Z) változóval, valamint különböző típusú hatásmechanizmusokkal (egy és többszörös hatás, egyszeres és többszörös függés, sajátdinamika és mellékhatás).....	98
4.3. ábra.	Példafeladat a PISA 2012-es Problémamegoldás moduljából (számítógép-alapú) .....	99
4.4. ábra.	Példafeladat a PISA 2015-ös Problémamegoldás moduljából (számítógép-alapú) .....	100
4.5. ábra.	Példafeladat a problémamegoldó feladatlapról .....	104
4.6. ábra.	A problémamegoldó képesség fejlettségi szintjének mérésére alkalmas feladatbank 100 itemének nehézségi indexe a 3–11. évfolyamos diákok	

	képességszintjének függvényében (minden 'x' 39 diákot reprezentál; a színek a feladatok típusát mutatják; 1. típus – iskolai tudás alkalmazását kívánó jóldefiniált problémák: kék, 2. típus – iskolai tudás alkalmazását kívánó rosszuldefiniált problémák: zöld, 3. típus – iskolán kívül tanultak alkalmazását kívánó rosszuldefiniált problémák: piros) .....	106
4.7. ábra.	A problémamegoldó képesség fejlődése .....	107
4.8. ábra.	A problémamegoldó képesség fejlődése 3–12. évfolyamon .....	108
4.9. ábra.	A problémamegoldó képesség fejlődése iskolatípusonkénti bontásban .....	109
4.10. ábra.	Különböző típusú problémák megoldásának személy/ item-térképe (1. dimenzió: iskolai tudás alkalmazását kívánó jóldefiniált problémák (kék); 2. dimenzió: iskolai tudás alkalmazását kívánó rosszuldefiniált problémák (zöld); 3. dimenzió: iskolán kívül tanultak alkalmazását kívánó rosszuldefiniált problémák (piros); minden 'x' 198 diákot reprezentál) .....	110
4.11. ábra.	Különböző típusú problémákon mutatott képességfejlődés (1. típus: iskolai tudás alkalmazását kívánó jóldefiniált problémák; 2. típus: iskolai tudás alkalmazását kívánó rosszuldefiniált problémák; 3. típus: iskolán kívül tanultak alkalmazását kívánó rosszuldefiniált problémák) .....	111
4.12. ábra.	Az eDia platformon a teszt egyik problémájának első része. (Egy hatalmas pillangóházban háromféle pillangót tenyésztasz: Pirospillét, Kékpillét és Zöldpillét. A pillangók sajnos nem úgy fejlődnek, ahogy te szeretnéd, ezért új virágokat telepítesz a pillangóházba: vadbogyót, sápadtlevelet és napfűvet. Ezek nektárja remélhetőleg segíti a pillangók fejlődését. Találd ki, hogy a vadbogyó, a sápadtlevél és a napfű nektárja milyen hatással van a különböző pillangófélék fejlődésére!) .....	120
4.13. ábra.	A TAO platformon a teszt egyik problémájának második része .....	121
4.14. ábra.	A kutatás felépítése – a problémák horgonyzása .....	122
4.15. ábra.	A dinamikus problémamegoldó képesség kétdimenziós (kétfaktoros) modellje (DPS: dinamikus problémamegoldó képesség) .....	127
4.16. ábra.	Az itemek nehézségi indexei a diákok és az évfolyamok átlagos képességszintjének függvényében (a minta átlagos képességszintje 0-ra transzformált, minden egyes 'X' nyolc diákot reprezentál; részletesebben l. Molnár, 2013; 2013-as adatfelvétel alapján) .....	132
4.17. ábra.	Az item-személy térkép az évfolyamok átlagos képességszintjének függvényében (minden egyes 'x' 8 diákot reprezentál) .....	133
4.18. ábra.	A 2013-as és 2014-es felmérés azonos tartalmú, de különböző feltételek mellett kiközvetített feladatai nehézségi szintjének alakulás egymás viszonylatában (minden egyes 'x' 8 diákot reprezentál) .....	134
4.19. ábra.	A négy különböző szcenárióban megvalósuló adatfelvétel közös itemeinek itemnehézségi mutatója .....	136
5.1. ábra.	Az interaktív problémákon mutatott problémamegoldó képesség fejlődése 3-12. évfolyamon (2014-es adatfelvétel alapján) .....	139
5.2. ábra.	A dinamikus problémamegoldó képesség fejlettségi szintjének eloszlása 3-9. évfolyamon .....	141

5.3. ábra.	A dinamikus problémamegoldó képesség teszt kétdimenziós személy/item térképe (A 2014-es és 2015-ös adatfelvétel egyesített adatbázisa alapján; minden egyes 'X' 35 diákot reprezentál.).....	142
5.4. ábra.	A dinamikus problémamegoldó képesség két dinemziójának, faktorának (tudás elsajátítás és tudás alkalmazás) fejlődési görbéje (egy.=egyetem 1. évfolyam).....	143
5.5. ábra.	Az intelligencia és a problémamegoldó gondolkodás fejlettségének kapcsolata (minden koeficiens $p < 0,01$ szinten szignifikáns; CFT: intelligencia, SPK: problémamegoldó képesség – statikus problémák, DPK: problémamegoldó képesség – dinamikus problémák; a vonal mintázata a kapcsolat erősségét mutatja: szaggatott vonal: leggyengébb, vastag folytonos vonal legerősebb kapcsolat. A lineáris korreláció egyenes vonallal, a parciális korreláció görbe vonallal ábrázolt.) .....	145
5.6. ábra.	Az induktív gondolkodás és a problémamegoldó képesség különböző típusú problémahelyzetekben mutatott fejlettségi szintje közötti kapcsolat (minden egyes együttható $p < 0,01$ szinten szignifikáns; SPK: statikus környezetben mutatott problémamegoldó képesség, DPK: dinamikus környezetben mutatott problémamegoldó képesség, IG: induktív gondolkodás; a kapcsolat erősségét jellemző vonalak magyarázatát l. 5.3. ábra).....	147
5.7. ábra.	Az induktív gondolkodás és a különböző problémahelyzetekben alkalmazott problémamegoldó képesség fejlettségi szintje kapcsolatának változása 5., 7. és 11. évfolyamon (a vonalak vastagsága a kapcsolat erősségét mutatja, l. 5.3. ábra magyarázata; n.s.: nem szignifikáns kapcsolat) .....	148
5.8. ábra.	Az induktív gondolkodás, az intelligencia és a különböző helyzetekben működtetett problémamegoldó képesség kapcsolata (minden együttható $p < 0,01$ szinten szignifikáns; a vonalak és rövidítések magyarázatát l. 5.3. és 5.4. ábra).....	150
5.9. ábra.	A DPK, az OKM matematika és szövegértés tesztjén elért teljesítmények kapcsolata (minden együttható $p < 0,01$ szinten szignifikáns; lineáris korreláció: folytonos vonal, parciális korreláció: szaggatott vonat).....	151
5.10. ábra.	A szülők iskolai végzettségének, a diákok induktív gondolkodása és intelligenciája fejlettségi szintjének előrejelző ereje a problémamegoldó képesség fejlettségi szintjére .....	154
5.11. ábra.	Az Országos kompetenciamérés, az iskolai sikeresség és a demográfiai háttérváltozók előrejelző ereje a diákok problémamegoldó képességszintjére – útelemzés (minden útvonal-együttható $p < 0,01$ szinten szignifikáns) .....	155
5.12. ábra.	Az első évfolyamos DIFER és induktív gondolkodás teszten nyújtott teljesítmények előrejelző ereje a hetedik évfolyamos problémamegoldó képességeszten nyújtott teljesítményre (minden útvonal-együttható $p < 0,01$ szinten szignifikáns) .....	156

5.13. ábra.	Az első évfolyamos DIFER, induktív gondolkodás és számolási készség teszten nyújtott teljesítmények előrejelző ereje a hetedik évfolyamos problémamegoldó képességeszten nyújtott teljesítményre .....	157
5.14. ábra.	A hetedik évfolyamos olvasás, induktív gondolkodás, matematika és angol teszteredmények előrejelző ereje a diákok problémamegoldó képességeszten nyújtott teljesítményre – útelemzés.....	158
5.15. ábra.	A hetedik évfolyamos diákok nemének, induktív gondolkodásuk fejlettségi szintjének, OKM matematika tesztjén elért eredményének és angol tudásszintjének előrejelző ereje a DPK teszten nyújtott teljesítményre – útelemzés (a modellben szereplő együtthatók $p < 0,01$ szinten szignifikánsak; IG: induktív gondolkodás; DPK: dinamikus problémamegoldó képesség; OKMm: az Országos kompetenciamérés matematika eredménye).....	159
5.16. ábra.	A MicroDYN problémák első fázisának feltérképezése (2 bemeneti, 2 kimeneti változó) .....	163
5.17. ábra.	A MicroDYN problémák első fázisának feltérképezése (2 bemeneti, 1 kimeneti változó) .....	164
5.18. ábra.	A legegyszerűbb, kettő bemeneti és egy kimeneti változót tartalmazó problémák feltérképezése során alkalmazott minimális stratégiák a teljesítmény függvényében (n.e.s: a szükséges információk kinyeréséhez nem elegendő stratégia; stratégiakódoláshoz l. 5.4.1. részt).....	170
5.19. ábra.	A két bemeneti és két kimeneti változót tartalmazó problémák feltérképezése során alkalmazott minimális stratégiák a teljesítmény függvényében (n.e.s: a szükséges információk kinyeréséhez nem elegendő stratégia; stratégiakódoláshoz l. 5.4.1. részt).....	171
5.20. ábra.	A három bemeneti és két kimeneti változót tartalmazó problémák feltérképezése során alkalmazott helyes VOTAT és nem VOTAT minimális stratégiák a teljesítmény függvényében.....	172
5.21. ábra.	A három bemeneti és három kimeneti változót tartalmazó problémák feltérképezése során alkalmazott helyes VOTAT és nem VOTAT alapstratégiák a teljesítmény függvényében .....	173
5.22. ábra.	A problémamegoldó stratégiákra alapuló látens profil elemzés (Jelmagyarázatban lévő feliratok: Class 5: A legegyszerűbb rendszereket sem kezelő problémamegoldók; Class 6: A legegyszerűbb rendszereket alacsony fokon átlátó problémamegoldók; Class 4: Egyszerű problémákat kisebb sikerekkel átlátó problémamegoldók; Class 2: A gyorsan tanuló problémamegoldók; Class 1: Az egyszerű rendszereket jól átlátó, de a bonyolultabbnál alulteljesítő problémamegoldók; Class 3: Magas szintű stratégiahasználók, problémamegoldók) .....	177
5.23. ábra.	A problémamegoldó stratégiákra alapuló látens profil elemzés (Jelmagyarázatban lévő feliratok: Class 5: A legegyszerűbb rendszereket sem kezelő problémamegoldók; Class 3: A legegyszerűbb rendszereket alacsony fokon átlátó problémamegoldók; Class 6: Egyszerű problémákat kisebb sikerekkel átlátó problémamegoldók; Class 2: A gyorsan tanuló	

problémamegoldók; Class 1: Az egyszerű rendszereket jól átlátó, de a bonyolultabbnál alulteljesítő problémamegoldók; Class 3: Magas szintű stratégiahasználók, problémamegoldók) .....	179
5.24. ábra. A problémamegoldó stratégiákra alapuló hét látens profilt definiáló elemzés csoportjai .....	180
5.25. ábra. Egyetemisták problémamegoldó stratégiáinak látens profilelemzése .....	181



## A DISSZERTÁCIÓBAN KÖZVETLENÜL FELHASZNÁLT SAJÁT KÖZLEMÉNYEK JEGYZÉKE

- Csapó Benő, Molnár Gyöngyvér és Nagy József (2015): A DIFER tesztek online változatával végzett mérések tapasztalatai. In: Csapó Benő és Zsolnai Anikó (szerk.): *Online diagnosztikus mérések az iskola kezdő szakaszában*. Oktatókutató és Fejlesztő Intézet, Budapest. 163–182.
- Csapó Benő, Molnár Gyöngyvér és R. Tóth Krisztina (2008): A papíralapú tesztek a számítógépes adaptív tesztelésig: a pedagógiai mérés-értékelés technikájának fejlődési tendenciái. *Iskolakultúra*, **18**. 3-4. sz. 3–16.
- Csapó, B., Molnár, Gy. és Nagy, J. (2014): Computer-based assessment of school readiness and early reasoning. *Journal of Educational Psychology*, **106**. 2. sz. 639–650.
- Molnár Gyöngyvér (2004): Hátrányos helyzetű diákok problémamegoldó gondolkodásának fejlettsége. *Magyar Pedagógia*, **104**. 3. sz. 319–337.
- Molnár Gyöngyvér (2005): Az objektív mérés megvalósításának lehetősége: a Rasch-modell. *Iskolakultúra*, **15**. 3. sz. 71–80.
- Molnár Gyöngyvér (2006): A tudáskonceptió változása és annak megjelenése a PISA 2003 vizsgálat komplex problémamegoldás moduljában. *Új Pedagógiai Szemle*, **56**. 1. sz. 75–86.
- Molnár Gyöngyvér (2006): A Rasch-modell alkalmazása a társadalomtudományi kutatásokban. *Iskolakultúra*, **16**. 12. sz. 99–113.
- Molnár Gyöngyvér (2006): *Tudástranszfer és komplex problémamegoldás*. Műszaki Kiadó, Budapest.
- Molnár Gyöngyvér (2007): Hátrányos helyzetű diákok problémamegoldó gondolkodásának longitudinális követése. *Magyar Pedagógia*, **107**. 4. sz. 277–293.
- Molnár Gyöngyvér (2010): Papír- és számítógép-alapú tesztelés összehasonlító vizsgálata problémamegoldó környezetben. In: Perjés István és Kozma Tamás: *Új Kutatások a Neveléstudományokban*. Aula Kiadó, Corvinus Egyetem, Budapest. 135–144.
- Molnár Gyöngyvér (2010): Technológiaalapú mérés-értékelés hazai és nemzetközi implementációi. *Iskolakultúra*, **20**. 7–8. sz. 22–34.
- Molnár Gyöngyvér (2011): Az információs-kommunikációs technológiák hatása a tanulásra és oktatásra. *Magyar Tudomány*, **172**. 9. sz. 1038–1047.
- Molnár Gyöngyvér (2011): Számítógépes játék-alapú képességfejlesztés: egy pilot vizsgálat eredményei. *Iskolakultúra*, **21**. 6–7. sz. 3–11.
- Molnár Gyöngyvér (2012): A problémamegoldó gondolkodás fejlődése: az intelligencia és szocioökonómiai háttér befolyásoló hatása 3-11. évfolyamon. *Magyar Pedagógia*, **112**. 1. sz. 41–58.
- Molnár Gyöngyvér (2013): *A Rasch modell alkalmazási lehetőségei az empirikus kutatások gyakorlatában*. Gondolat Kiadó, Budapest.
- Molnár Gyöngyvér (2013): Mindennapi helyzetekben alkalmazott problémamegoldó stratégiák változása. *Iskolakultúra*, **23**. 7–8. sz. 31–43.
- Molnár Gyöngyvér (2013): Terület-specifikus komplex problémamegoldó gondolkodás fejlődése. In: Molnár Gyöngyvér és Korom Erzsébet (szerk.): *Az iskolai*

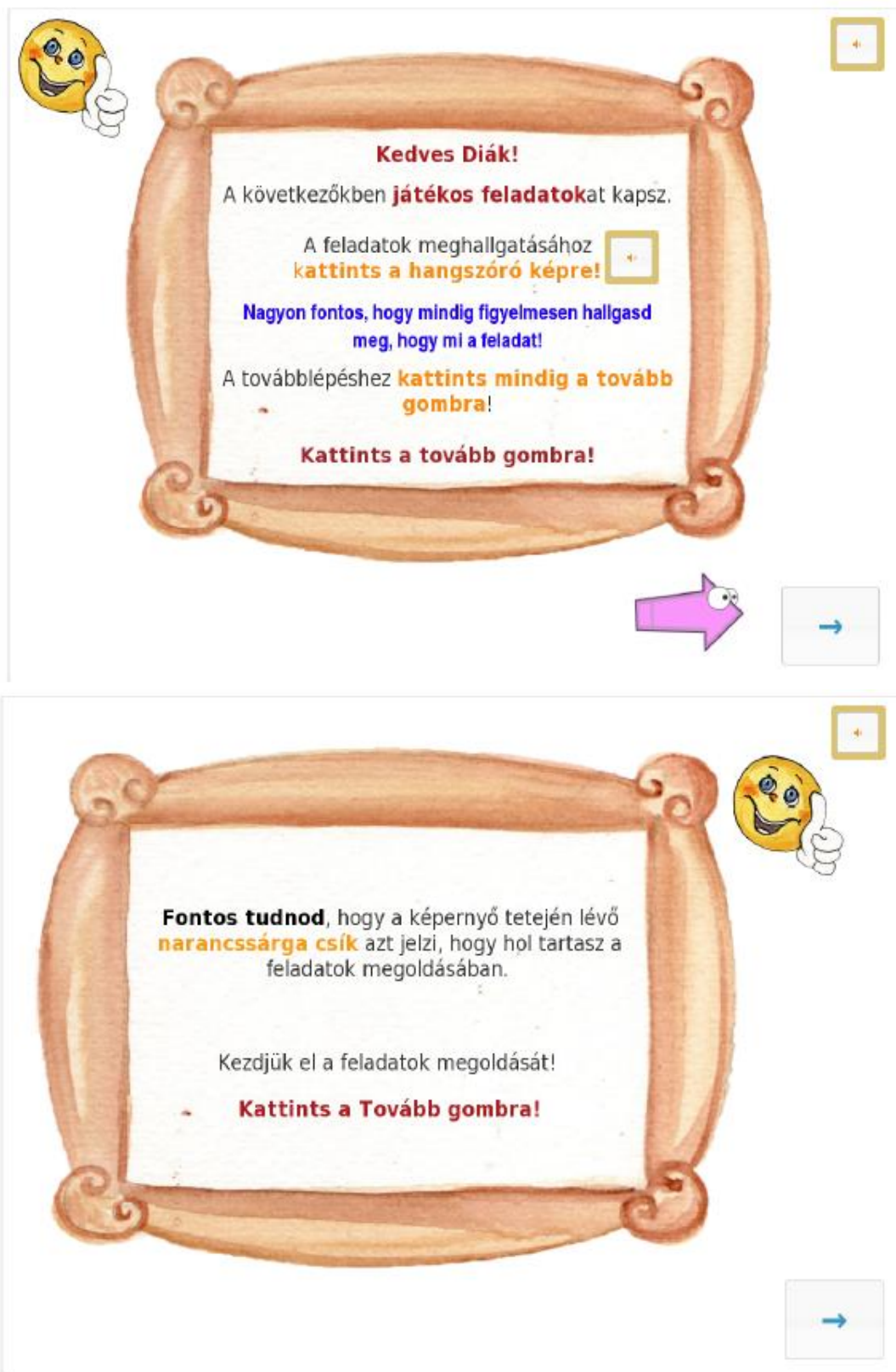
- sikerességet befolyásoló kognitív és affektív tényezők értékelése.* Nemzedékek Tudása Tankönyvkiadó, Budapest. 161–180.
- Molnár Gyöngyvér (2015): Az óvoda és iskola feladatai az értelmi képességek fejlesztése terén. In: Kónyáné Tóth Mária és Molnár Csaba (szerk.): *Tartalmi és szervezeti változások a köznevelésben.* Suliszerviz Oktatási és Szakértői Iroda, Suliszerviz Pedagógiai Intézet, Debrecen. 179–190.
- Molnár Gyöngyvér és Csapó Benő (2003): A képességek fejlődésének logisztikus modellje. *Iskolakultúra*, **13.** 2. sz. 57–69.
- Molnár Gyöngyvér és Csapó Benő (2013): Az eDia online diagnosztikus mérési rendszer. XI. Pedagógiai Értékelési Konferencia. Szeged, 2012. április 11-13. 82.
- Molnár Gyöngyvér és Magyar Andrea (2015): A számítógép alapú tesztelés elfogadottsága pedagógusok és diákok körében. *Magyar Pedagógia*, **115.** 1. sz. 49–66.
- Molnár Gyöngyvér és Pásztor Attila (2015): A számítógép alapú mérések megvalósíthatósága kisiskolás diákok körében: első évfolyamos diákok egér- és billentyűzet-használati képességeinek fejlettségi szintje. *Magyar Pedagógia*, **115.** 3. sz. 237–252.
- Molnár Gyöngyvér és Pásztor-Kovács Anita (2015): A problémamegoldó képesség mérése online tesztkörnyezetben. In: Csapó Benő és Zsolnai Anikó (szerk.): *Online diagnosztikus mérések az iskola kezdő szakaszában.* Oktatókutató és Fejlesztő Intézet, Budapest. 279–300.
- Molnár Gyöngyvér és Pásztor-Kovács Anita (2015): A számítógépes vizsgáztatás infrastrukturális kérdései: az iskolák eszközparkjának helyzete és a változás tendenciái. *Iskolakultúra*, **25.** 4. sz. 49–61.
- Molnár Gyöngyvér, Tongori Ágota és Pluhár Zsuzsa (2015): Az informatikai műveltség online mérése. In: Csapó Benő és Zsolnai Anikó (szerk.): *Online diagnosztikus mérések az iskola kezdő szakaszában.* Oktatókutató és Fejlesztő Intézet, Budapest. 241–260.
- Molnár, Gy., Greiff, S., Csapó, B. (2013): Inductive reasoning, domain specific and complex problem solving: relations and development. *Thinking skills and Creativity*, **9.** 8. sz. 35–45.
- Molnár, Gy., Greiff, S., Wüstenberg, S. és Fischer, A. (2014): Empirical study of computer based assessment of domain-general dynamic problem solving skills. In: Csapó, B. Funke, J. és Schleicher, A. (2016, szerk.): *On the nature of problem solving.* OECD, Paris. 123–143.
- R. Tóth Krisztina, Molnár Gyöngyvér, Thibaud Latour és Csapó Benő (2011): Az online tesztelés lehetőségei és a TAO platform alkalmazása. *Új Pedagógiai Szemle*, **61.** 1–2–3–4–5. sz. 8–22.

## **8. MELLÉKLETEK**

## **Mellékletek jegyzéke**

8.1.	Géphasználat teszt – 1. évfolyam, részlet (2014)	236
8.2.	Dinamikus problémamegoldó képesség tesztek linkelése	247
8.3.	Dinamikus problémamegoldó képesség teszt – részlet	248

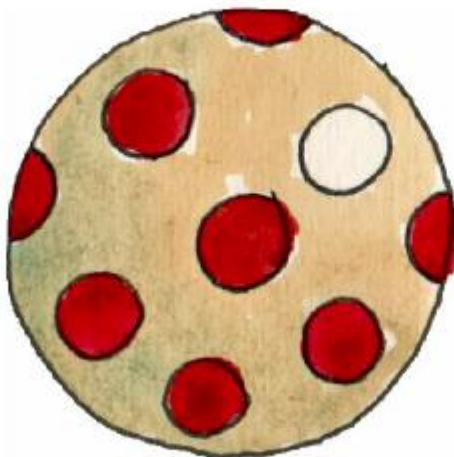
## 8.1. MELLÉKLET: GÉPHASZNÁLAT TESZT – 1. ÉVFOLYAM, RÉSZLET (2014)



Barátod fehér pöttyös helyett piros pöttyös labdát szeretne.  
Segíts neki befejezni a színezést!

**Kattints rá az utolsó fehér pöttyre, hogy az is pirosra változzon!**

Ha készen vagy, most is, és ezentúl mindig a nyílra kattintva léphetsz tovább.



Kilyukadt a vödör. Foltozd be minél gyorsabban!

**Kattints a lyukra, mielőtt letelik az idő!**



5



Pukkaszd ki a lufikat! **Kattints rájuk!**



Varázsolj piros orrot a bohócnak! **Kattints rá!**



**Pukkaszd ki az összes buborékot,** mielőtt lejár az idő!



8



Kapcsold be a rádiót!

**Kattints a rádió gombjára!**





Rögzítsd a képet!

**Kattints a kép sarkaiban található szögekre!**



7

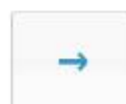


Csatlakoztasd a vagon a mozdonyhoz! **Kattints a vagonra, majd az egér bal gombját lenyomva tartva húzd rá a fekete körre!**



Mikulás már várja a manókat. Segíts az utolsónak felöltözni!

**Húzd rá a ruháját és a sapkáját!**



Már csak a névmatricák hiányoznak a füzeteidről.

**Húzd az egyes matricákat az alattuk található füzetekre!**



10



Elszakadt a gyöngysor. Fűzd fel gyorsan az elgurult gyöngyöket, mielőtt lejár az idő! **Húzd őket a vonalra!**



7



Színezd ki a virág szirmait **pirosra!**

**Kattints először a piros színre, majd a virág szirmaira!**



**Színezd ki a fehéren maradt szőlőszemeket**, mielőtt lejár az idő!



8



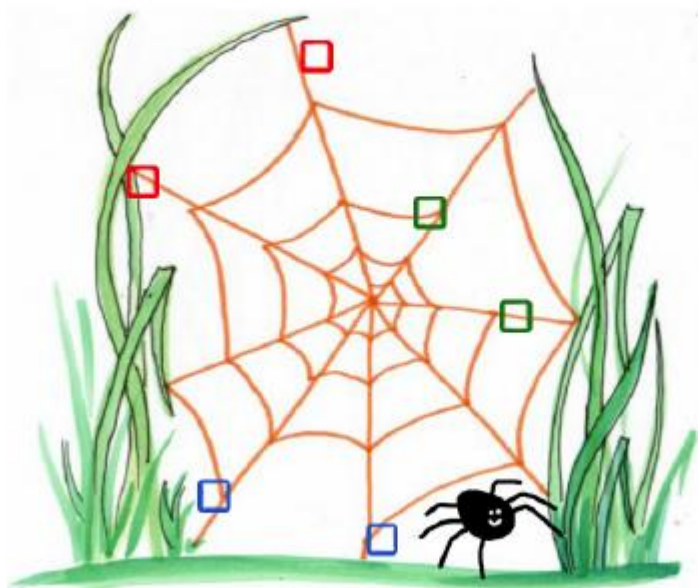
Segíts Katinak kihúzni a szárítókötelet a teregetéshez!

**Kattints az egyik, majd a másik a fára!**



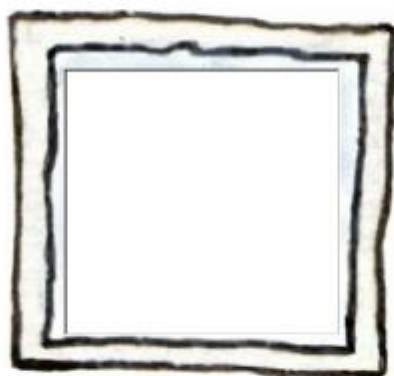
Szétszakadt a pókháló. Segíts a póknak a hiányzó szálakat megszőni!

**Kattints az azonos színű keretekre a pókháló javításához!**



**Írd rá az ablakra az 5-ös számot!**

Ehhez kattints az ablak közepére, keresd meg a billentyűzeten az 5-ös számot, és üsd le!

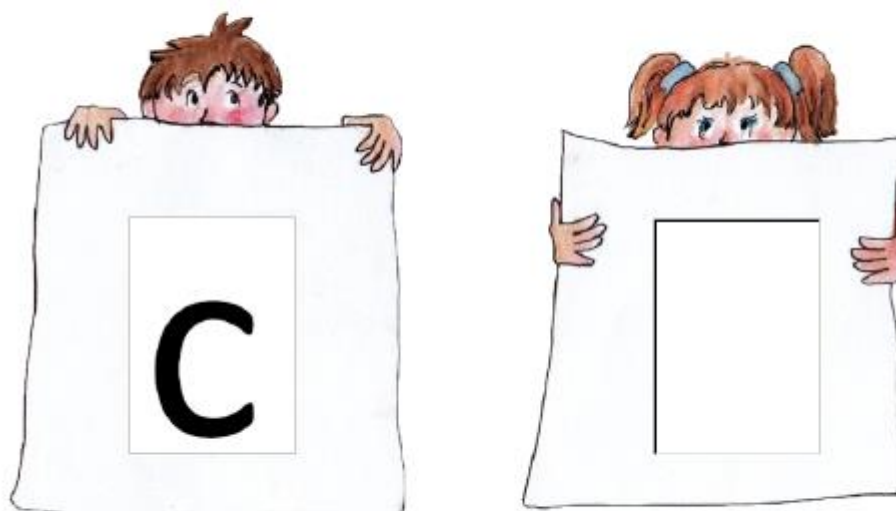


**Írd a házzámtáblákra a ház alatti számot!**

Először kattints a táblára, utána tudod beírni a házzámot.

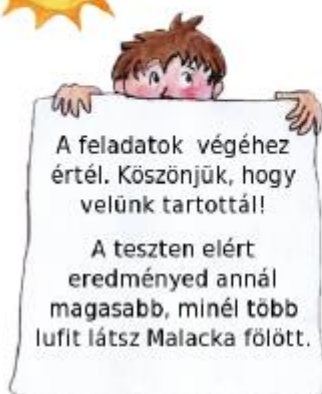
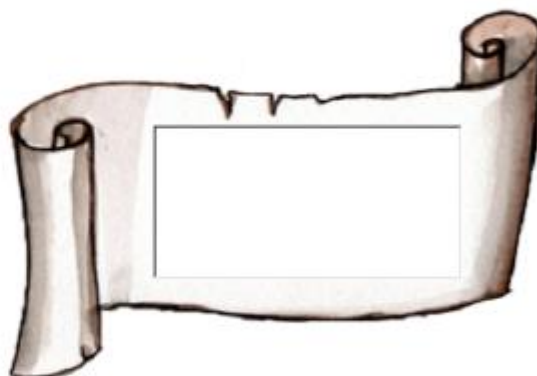
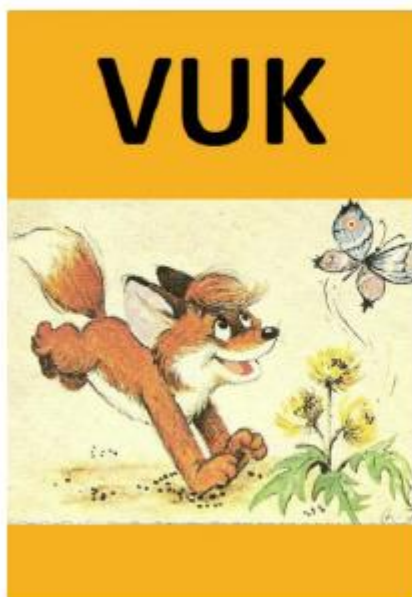


**Másold a kisfiú tábláján lévő betűt a kislány táblájára!**





Másold át a mesekönyv címét a lapra!



## 8.2. MELLÉKLET: DINAMIKUS PROBLÉMAMEGOLDÓ KÉPESSÉG TESZTEK LINKELÉSE

Név	Struktúra (be/kime- neti változók száma)	Saját- dina- mika	Évfolyam					
			3 (modell- ből nélkül)	4	5	6-8	9-12	egye- tem
Autó (példa)	1-1		+	+				
Sport (példa)	2-2				+	+		
Limonádé	2-1		+	+	+	+	+	+
Rajzóra	2-2		+	+	+	+		+
Korrepetálás	2-2		+	+	+			
Cica	2-2		+	+	+	+	+	+
Kismotor	3-2		+	+	+	+	+	+
Társasjáték	3-3		+	+	+	+	+	+
Parfüm	3-3						+	+
Kertimunka	3-2	+			+	+	+	+
Kézilabda	3-3				+	+	+	+
Világűr	3-2	+				+	+	+
Elsősegély	3-3	+				+	+	+
Pillangó	3-3	+					+	+
Üzem	3-3	+					+	

A különböző adatfelvételekben nem minden esetben ugyanazon teszt kiköszvetítése valósult meg azonos évfolyamon – a jelölt feladatokból történt az adott évfolyam kapcsán a válogatás.




## 8.3. MELLÉKLET: DINAMIKUS PROBLÉMAMEGOLDÓ KÉPESSÉG TESZT – RÉSZLET



A továbbiakban játékos feladatokat kapsz, ahol ki kell találnod, mi mivel van kapcsolatban.

A példafeladatban a sportolás és olvasás szeretetének kitartásra és erőre gyakorolt hatását kell felfedezned.

A következő oldalon lévő videó megmutatja, hogy tudod ezt megtenni.

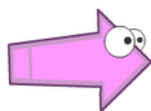
A videó elindításához kattints a 

alakú gombra. Ha megszeretnéd állítani a videót, kattints a



alakú gombra.

Kattints a tovább gombra!



Tovább ➤



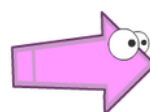
A példafeladatban láttad a feladatok megoldási módját.

A továbbiakban hasonló feladatokat fogsz kapni.

A feladatok első részében fel kell majd fedezned az elemek közötti összefüggéseket és azt berajzolni az ábra alatti modellbe.

A feladatok második részében az előre megadott célértékeket kell elérned.

Kattints a tovább gombra!



◀ Vissza

Tovább ➤



## A moped

Az új kismotoroddal meg szeretnéd látogatni a barátaidat. A kismotorba egyszerre háromféle üzemanyagot tudsz tankolni és ezek különböző keverékét használni: Carenol, Noresal és Farunin.

**Találd ki, hogy a keverék üzemanyag Carenol, Noresal és Farunin tartalma milyen hatással van a moped végsebességére és a kipufogógáz összetételére!**

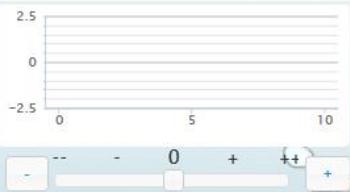
**Kattints a tovább gombra!**

**Tovább** ➔

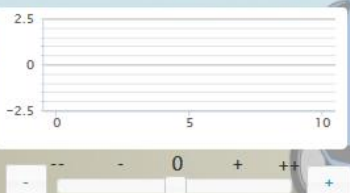


Találd meg az összefüggéseket, és sejtésed rajzold a modellbe!


**Carenol**



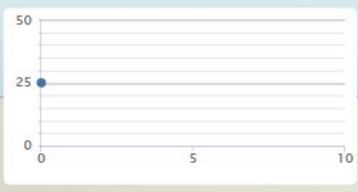
**Noresal**



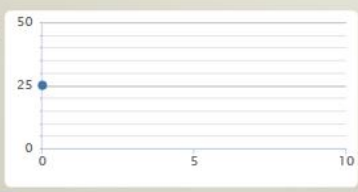
**Farunin**



**Végsebesség**



**Kipufogógáz összetétele**



179

Segítség
Törlés
Alkalmazás

Carenol

Noresal

Farunin


Végsebesség

Kipufogó gáz

Tovább ➔

Nem rajzoltál be egy nyilat sem a modellbe.  
Kérlek, lépj vissza az előző oldalra, és pótold!


Vissza




A következő oldalon megnézheted, hogy a valóságban  
hogyan befolyásolják a különböző üzemanyagok a  
kismotor végsebességét és a kipufogógáz összetételét.

**Hogyan adagoljuk a háromféle összetevőt, hogy  
maximum 4 lépésben elérjük a motor előre  
meghatározott végsebességét és a kipufogógáz  
összetételét?**

**Kattints a tovább gombra!**

Tovább 



Érd el az előre meghatározott céltartományt legfeljebb 4 lépésben!

**179**

**Carenol**

**Noresal**

**Farunin**

**Végsebesség**

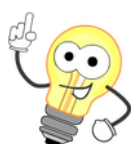
**Kipufogógáz összetétele**

**Segítség** **Törlés** **Alkalmazás**

**Tovább**

```

graph LR
    Carenol --> Végsebesség
    Noresal --> Végsebesség
    Farunin --> Kipufogógáz_összetétele
  
```



Befejezted a feladat megoldását!

Kattints a tovább gombra a következő feladathoz!

**Tovább**